PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/12, C07K 14/47 | A1

(11) International Publication Number: **WO 99/63080**

(43) International Publication Date: 9 December 1999 (09.12.99)

(21) International Application Number: PCT/EP99/03848

(22) International Filing Date: 2 June 1999 (02.06.99)

(30) Priority Data:
9811962.1 3 June 1998 (03.06.98) GB

(71) Applicant *(for all designated States except US)*: JANSSEN PHARMACEUTICA N.V. [BE/BE]; Turnhoutseweg 30, B–2340 Beerse (BE).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: LUYTEN, Walter, Herman, Maria, Louis [BE/BE]; Janssen Pharmaceutica N.V., Turnhoutseweg 30, B–2340 Beerse (BE). DE RAEYMAEKER, Marc, Carl [BE/BE]; Janssen Pharmaceutica N.V., Turnhoutseweg 30, B–2340 Beerse (BE). GEYSEN, Johan, Jozef, Gustave, Hendrik [BE/BE]; Janssen Pharmaceutica N.V., Turnhoutseweg 30, B–2340 Beerse (BE). BOGAERT, Thierry, A., O., E. [BE/BE]; Wolvendreef 26G, B–8500 Kortrijk (BE). MAERTEN, Luc, Jacques, Simon [BE/BE]; Devgen N.V., Blok DF 1.60.14, Technologiepark Zwijnaarde 9, B–9052 Zwijnaarde (BE). VERHASSELT, Peter [–/BE]; Janssen Pharmaceutica N.V., Turnhoutseweg 30, B–2340 Beerse (BE). VAN DE CRAEN, Marc [–/BE]; Devgen N.V., Blok DF 1.60.14, Technologiepark Zwijnaarde 9, B–9052 Zwijnaarde (BE).

(74) Agents: MAYES, Stuart, David et al.; Boult Wade Tennant, 27 Furnival Street, London EC4A 1PQ (GB).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: HUMAN HOMOLOGUE OF UNC–53 PROTEIN OF *C. ELEGANS*

(57) Abstract

There is disclosed human homologues of the UNC–53 protein of *C. elegans* and cDNA sequences coding for said homologues or functional equivalents thereof. The invention also relates to processes for identifying compounds which control cell behaviour, compounds identified and pharmaceutical compositions containing them in addition to processes and assays for identifying disease states in which said gene or protein is dysfunctional.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

HUMAN HOMOLOGUE OF UNC-53 PROTEIN OF *C. ELEGANS*

The present invention relates to a vertebrate homologue of UNC-53 protein of C. elegans and cDNA sequences coding for said homologues or functional equivalents thereof. The invention also relates to processes for identifying compounds which control cell behaviour, compounds identified and pharmaceutical compositions containing them in addition to processes and assays for identifying disease states in which said gene or protein is dysfunctional.

The control of cell motility, cell shape and directionality of cell outgrowth of axones or other cell outgrowths is an essential feature in the morphogenesis and function of both unicellular and multicellular organisms.

Some cell surface proteins and extra-cellular molecules controlling the directionality and potential of cell migration have been identified, although the processes involved are not generally understood. It is generally considered that a long-range migration of a cell process (also known as a growth cone extension) is a stepwise event, whereby prior to and after each extension there is the formation of a structure at the leading edge of the cell. Localised stabilisation of the actin cytoskeleton and association with plus end regions of microtubules is a general cell biological process underlying the choice of directional extension.

The present inventors have surprisingly found a new human gene/protein belonging to the UNC-53 family that binds microtubules and, in particular, the plus-end regions of microtubules.

A gene from the free-living nematode Caenorhabditis elegans designated "unc-53" has been previously identified and cloned (Abstract,

International C. elegans Meeting, June 1-5 1991, Madison, Wisconsin, 58, Bogaert and Goh). The present inventors previously identified UNC-53 protein as a signal transducer or signal integrator controlling the directionality of cell migration and/or cell shape in C. elegans (WO 96/38555).

The C. elegans UNC-53 protein (Ceunc53) and previously found human homologues thereof (hs-unc53/1 and hs-unc53/2) were found to encode a signal transducer or a signal integrator, controlling the directionality of a cell migration, cell shape and growth extension. Evidence indicates that the presently found homologue designated (hs-unc53/3) might act as an adapter linking extracellular signals to the actin cytoskeleton. Firstly hs-unc-53/3 shows homology to the cortical actin binding proteins, and the Ce-UNC-53 protein has been shown to bind F-actin in vitro and leads to actin re-organization in vivo when expressed in mammalian cells, leading to an increased number of filopodia and lammelipodia. Furthermore, increased neurite extension and increased cell motility could be observed. Hs-UNC-53-3 may play an important role in the development of various diseases.

According to a first aspect of the present invention there is provided a vertebrate protein homologue of an UNC-53 protein of C. elegans, which protein comprises an amino acid sequence having one or more of sequence blocks A, B, C, D, E, F, G or H as illustrated in figure 4 or which differs from said blocks in conservative amino acid changes.

According to a further aspect of the present invention, there is provided a vertebrate protein homologue of UNC-53 protein of C. elegans or a functional equivalent, derivative or bioprecursor thereof, having an amino acid sequence encoded by the nucleotide sequence illustrated in figure 1(e).

For the purposes of the present invention a "derivative" should be taken to mean mutational derivatives, fusions, internal deletions, splice variants and muteins.

Preferably, said vertebrate homologue is a human protein, and preferably a mammalian or a mouse protein.

A further aspect of the invention comprises a vertebrate homologue comprising an amino acid sequence as shown in figure 1(f) or the variants thereof or an amino acid sequence which differs from the amino acid sequences shown in figure 1(f) to a significant extent only in one or more conservative amino acid changes.

In a further aspect of the present invention there is also provided a nucleic acid molecule, which is preferably DNA, and which encodes a vertebrate homologue of UNC-53 protein of C. elegans, or a functional equivalent derivative, fragment or bioprecursor of said homologue according to the invention. Preferably, the cDNA comprises a sequence of nucleotides encoding an amino acid sequence as illustrated in figure 1(f) or the variants thereof or an amino acid which differs from the sequences shown in these figures to a significant extent only in one or more conservative amino acid changes. Preferably the DNA is cDNA, which cDNA comprises the sequence shown in figure 1(e)or the variants indicated therein. Also provided by the present invention is a nucleic acid sequence capable of hybridising to the nucleic acid or DNA sequences according to the invention under high stringency conditions, which conditions are well known to those skilled in the art.

The cDNA according to the invention may be included in an expression vector which may itself be used to transform or transfect a host cell, which cell may be bacterial or eukaryotic in origin including such as, for example an animal or plant cell a fungal

cell or an insect cell. Thus, advantageously, once the cDNA corresponding to the genome of the vertebrate homologue of UNC-53 of C. elegans according to the invention is synthesised, using for example, reverse transcriptase or the like, a range of cells, tissues or organisms may be transfected following incorporation of the selected cDNA clone into an appropriate expression vector. The expression vector according to the invention may comprise a promoter of C. elegans or one of human, mouse or viral origin and optionally a sequence encoding a reporter molecule, such as, for example, green fluorescent protein.

The present invention, therefore, also further comprises a transgenic cell, tissue or organism comprising a transgene capable of expressing a vertebrate homologue of UNC-53 protein of C. elegans according to the invention. The term "transgene capable of expressing a vertebrate homologue of UNC-53 protein of C. elegans" as used herein means a suitable nucleic acid sequence which leads to the expression of a vertebrate homologue of UNC-53 protein of C. elegans according to the invention having the same function and/or activity. The transgene may include, for example, genomic nucleic acid isolated from the appropriate vertebrate or synthetic nucleic acid including cDNA. The term "transgenic organisms, tissues or cells, as used herein means any suitable organism and/or part of an organism, tissue or cell, that contains exogenous nucleic acid either stably integrated in the genome or in an extrachromosomal state.

Preferably the transgenic cell comprises any of, a COS cell, HepG2 cell, MCF-7 or N4 neuroblastoma cell, a NIH3T3 cell, a colorectal or carcinoma cell or a human derived cell such as a fibroblast or the like. The transgenic organism may be an insect, a non-human animal or a plant and preferably C. elegans or a

related nematode. Preferably, the transgene comprises the nucleic acid or cDNA sequence encoding the vertebrate homologue according to the invention as described above. The transgene preferably comprises an expression vector according to the invention.

The term "functional fragment" as used herein should be taken to mean a fragment of the gene coding for the vertebrate homologue of the UNC-53 protein of C. elegans according to the invention. For example, the gene may comprise deletions or mutations but may still encode a functional vertebrate homologue of UNC-53 protein.

Further provided by the present invention is a method of producing a mutant vertebrate non-human organism having a mutation in the wild-type gene coding for the vertebrate homologue of UNC-53 protein according to the invention, which mutation affects cell behaviour or the regulation of cell motility or the shape or the direction of cell migration or microtubule plus end stability or function and localisation of protein complexes located thereon, which method comprises inducing a mutation in the vertebrate homologue of UNC-53 protein in said organism. These mutant organisms may be used in a screen to identify the effects of compounds on these cell functions.

The vertebrate homologue of UNC-53 protein of C. elegans or the cDNA or genomic DNA encoding it or a functional equivalent, derivative, fragment or bioprecursor of said homologue, may advantageously be used as a medicament, or in the preparation of a medicament to treat or prevent disorders associated with inhibition of overexpression of the vertebrate homologue of UNC -53 according to the invention. Such disorders may be alleviated by promoting neuronal regeneration, revascularisation or wound healing or the treatment of chronic neurodegenerative disorders,

psychiatric disorders or acute traumatic injuries or fibrotic disease or disease in which physiological events requiring the polarity of cells or epithelia are abnormally functioning. Accordingly, the vertebrate homologue according to the invention, dominant positive or negative mutants thereof, or inhibitors thereof may advantageously be used to induce or alleviate contact inhibition in a cell or in preventing carcinoma development. Typically, the above medical conditions may be treated in mammals and more preferably humans by either the homologue of UNC-53 protein or alternatively by a nucleic acid coding for the protein or the protein itself according to the invention. Alternatively an antisense oligonucleotide to said UNC-53 vertebrate homologue may be used to prevent it's expression. Examples of other nucleic acid sequences which may be used include 3' untranslated regions of mRNA which could be used to prevent transcription of the genomic sequence encoding for the vertebrate homologue of UNC-53 protein according to the invention.

The vertebrate homologue of UNC-53 protein according to the invention may be incorporated into a pharmaceutically acceptable composition together with a suitable carrier, diluent or excipient therefor. The pharmaceutical composition may advantageously comprise, additionally or alternatively, the nucleic acid sequence according to the invention as defined above.

The induction or inhibition of the expression of hu-UNC-53/3 by pharmacological means may advantageously be used to induce neuronal regeneration, revascularisation or wound healing or be involved in the treatment of chronical neurodegenerative disorders, or acute traumatic injuries or fibrotic diseases, or physiological events requiring the polarity of cells, or oncology and

metastasis of cells, or apoptotic pathways.

The present invention therefore also provides for a method of determining whether a compound is an inhibitor or enhancer of the regulation of cell behaviour, growth, transformation, cell shape or motility or the direction of cell migration, microtubule plus end stability or function and localisation of protein complexes thereon, which method comprises contacting said compound with a transgenic cell according to the invention and screening for a phenotypic change in said cell. The method can therefore be used to determine whether the compound comprises an inhibitor or an enhancer of the signal transduction pathway of said transgenic cell of which pathway said vertebrate homologue of UNC-53 protein according to the invention is a component, or whether said compound is an inhibitor or an enhancer of a parallel or redundant signal transduction pathway in said cell. The present invention also provides a method to determine that the protein in said signal transduction pathway is a vertebrate homologue of UNC-53 protein of C. elegans according to the invention.

Preferably, the phenotypic change to be screened comprises a change in cell shape or a change in cell motility. Where a transgenic cell is used in accordance with one embodiment of the method of the invention, an N4 neuroblastoma cell may be used and in such an embodiment the phenotypic change to be screened may be the length of neurite growth, changes in filopodia outgrowth, changes in ruffling behaviour or cell adhesion, any change in microtubule cytoskeleton, any change in localisation of proteins on plus end regions of microtubules or any change in a cell such as apoptosis. In an alternative embodiment of the method of the invention, the transgenic cell may comprise an MCF-7 breast carcinoma cell. Typically in such an embodiment the phenotypic change

to be screened comprises the extent of phagokinesis or filopodia formation. In an alternative embodiment of this aspect of the invention, the transgenic cell may comprise an NIH3T3 cell. Typically in such an embodiment the phenotypic change to be screened comprises loss of contact inhibition of foci formation. The method according to the invention, may also utilise a mutant cell or mutant organism according to the invention as described above, where the mutant cell is capable of growing in tissue culture or in vivo and either of which cell or organism has a mutation in the wild-type unc-53 gene.

In accordance with the present invention, a "phenotypic change", may comprise any phenotype resulting from changes at any suitable point in the life cycle of the cell, tissue or organism defined above, which change can be attributed to the expression of the transgene of the invention such as for example, growth, viability, morphology, behaviour, movement, cell migration or cell process or growth cone extension of cells and includes changes in body shape, locomotion, chemotaxis, contact inhibition, mating behaviour or the like. The phenotypic change may preferably be monitored directly by visual inspection of the cell as a whole or by monitoring the F-actin cytoskeleton microtubule network and plus end stability of microtubules or proteins thereon or alternatively by for example measuring indicators of viability including endogenous or transgenically introduced histochemical markers or other reporter genes, such as for example β-galactosidase or green fluorescent protein.

A compound which is identifiable by the method according to the invention as described above, as an enhancer of the processes identified above such as the regulation of cell shape or motility or the direction of cell migration may be used as a medicament, or

alternatively in the preparation of a medicament, for promoting neuronal regeneration, revascularisation or wound healing, or for treatment of chronic neuro-degenerative diseases or acute traumatic injuries or fibrotic disease. Examples of promoting neuronal regeneration include, for example, peripheral nerve regeneration after trauma and spinal cord trauma.

Where a compound is identified in accordance with the method described above as being an inhibitor of the regulation of cell shape or mobility or the direction of cell migration, the compound may be used as a medicament, or in the preparation of a medicament, for substantially alleviating spread of disease inducing cells, such as in spread of carcinoma, or the like in metastasis or in alleviating loss of contact inhibition. Advantageously, any of the compounds which may have been identified as an inhibitor or an enhancer in accordance with the method as described above, may also be included in a pharmaceutical composition comprising the respective compound and a pharmaceutically acceptable carrier, diluent or excipient therefor.

The particular mechanism of action of a compound identified as either an inhibitor or an enhancer of the cell motility shape, growth or direction of cell migration or microtubule association or to the plus end region thereof is not limiting. Preferably the compound acts as an inhibitor or enhancer of a signal transduction pathway. The compound may also act on a parallel pathway or directly on the vertebrate homologue of UNC-53 protein of C. elegans. For example, the method of action of the compound may include direct interaction with the vertebrate homologue of UNC-53 protein, interaction with processes for regulating phosphorylation or dephosphorylation of the vertebrate homologue of UNC-53 or with processes regulating activity of an unc-53

gene or with processes for post-transcriptional or post-translational modification or the like.

Preferably the compound is identified by the method according to the invention as an inhibitor or an enhancer, by utilising differences of phenotype of the cell, tissue or organism, which are visible to the eye. Alternatively indicators of viability including endogenous or transgenically introduced histochemical markers or a reporter gene may be used.

According to a further aspect of the invention there is also provided a transgenic cell or tissue culture which has been constructed to comprise a promoter sequence of a gene coding for a vertebrate homologue of UNC-53 of C. elegans according to the invention operably linked to a nucleic acid sequence encoding a reporter molecule. Preferably, the reporter sequence encodes for a detectable protein, for example one which may be monitored by eye inspection such as antibiotic resistance, β-galactosidase or a molecule detectable by spectrophotometric, spectrofluorometric, luminescent or radioactive assays.

The present invention also provides a method of determining whether a compound is an inhibitor or an enhancer of transcription of a gene coding for a vertebrate homologue of UNC-53 protein in C. elegans, according to the invention which method comprises the steps of:

(a) contacting said compound with a transgenic cell according to the invention as described above,

(b) monitoring the level of said reporter molecule and comparing results obtained from this monitoring step with a control comprising a transgenic cell having the promoter sequence of a gene coding for a vertebrate homologue of UNC-53 protein, or a functional fragment of said

homologue and the reporter molecule, in the absence of the compound.

In one embodiment of the method according to this aspect of the invention the reporter molecule may comprise messenger RNA.

A compound identified as an enhancer of transcription of the gene coding for the vertebrate homologue of UNC-53 protein of C. elegans or a functional equivalent, derivative or bioprecursor of said homologue may also be used as a medicament, or in the preparation of a medicament, for promoting neuronal regeneration, revascularisation or wound healing, or for treatment of chronic neuro-degenerative diseases or acute traumatic injuries or fibrotic disease. Furthermore, such compounds may be included in a pharmaceutical composition including a pharmaceutically acceptable carrier, diluent or excipient therefor. Any compounds identified as inhibitors of transcription may, advantageously, be used in alleviating the spread of disease inducing cells such as carcinomas or metastasis or loss of contact inhibition.

The present invention also provides a kit for determining whether a compound is an enhancer or an inhibitor of the regulation of cell growth, transformation, cell motility or shape or the direction of cell migration which kit comprises at least one transgenic or mutant cell or transgenic or mutant non-human organism according to the invention as described above and a plurality of wild-type cells or a wild-type organism of the same type, or a cell line or tissue culture and means for contacting said compound with said cell or organism.

Also provided by the present invention is a kit for determining whether a compound is an inhibitor or an enhancer of transcription of a gene coding for a vertebrate homologue of UNC-53 protein of C. elegans

according to the invention which kit comprises at least one transgenic cell or cells according to the invention, means for contacting said compounds with said cells and means for monitoring the level of transcription of said transgenic cell or cells according to the invention.

For the purposes of the present invention, the term "gene coding for a vertebrate homologue of UNC-53 or a functional fragment of said homologue" includes the nucleic acid sequence shown in figure 1 or a fragment thereof, including the differentially spliced isoforms and transcriptional starts of the nucleic acid sequence and which sequence encodes a vertebrate homologue of UNC-53 protein or a functional equivalent, derivative, fragment or bioprecursor of the protein.

The present invention also provides methods of identifying genes of vertebrates or fragments of said genes, which encode proteins which are active in the signal transduction pathway of which the vertebrate homologue of UNC-53 according to the present invention is a component. A preferred method comprises hybridizing to an appropriate cDNA library a nucleotide sequence, as defined herein, or a fragment thereof under appropriate conditions of stringency in order to identify genes having statistically significant homology with the cDNA clones of any one of the cDNA sequences according to the invention described above.

Furthermore, there is also provided by the present invention a method of identifying a protein which is active in the signal transduction pathway of a cell of which a vertebrate homologue of UNC-53 protein of C. elegans according to the invention is a component. According to this aspect of the invention, the method comprises;

(a) contacting an extract of said cell with an

antibody to the vertebrate homologue of UNC-53 protein or a functional equivalent, fragment or bioprecursor of said protein,

(b) identifying the antibody/vertebrate homologue of UNC-53 complex, and

(c) analysing the complex to identify any protein bound to the vertebrate homologue of UNC-53 protein other than the antibody.

The vertebrate homologue of UNC-53 protein, therefore may bind regions of other proteins involved in the signal transduction pathway. It is also possible to sequentially identify a whole range of proteins involved in the signal transduction pathway.

Antibodies to the vertebrate homologue of UNC-53 protein may be produced according to known techniques as would be known to those skilled in the art. For example, polyclonal antibodies may be prepared by inoculating a host animal, such as a mouse, with a protein or epitope of a protein according to the invention and recovering immune serum.

This aspect of the invention, further comprises a method of identifying a further protein or proteins which are active in the signal transduction pathway of a cell of which the vertebrate homologue of UNC-53 is a component which method comprises:

(a) forming an antibody to the first identified protein bound to the vertebrate homologue of UNC-53 protein in the method as described above,

(b) contacting a cell extract with the antibody,

(c) identifying any antibody/protein complex,

(d) analysing the complex to identify any further protein bound to the first protein other than the antibody, and

(e) optionally repeating steps (a) to (d) to identify further proteins in the pathway.

According to this aspect of the present invention, the antibody starts the process by binding

to the vertebrate homologue of UNC-53 protein according to the invention in the signal transduction or oncogenic pathways. Any other proteins found complexed to the bound antibody or UNC-53 protein can then be used to identify further interacting proteins involved in the pathway.

It may also be possible to identify proteins involved in the signal transduction pathway of a cell of which the vertebrate homologue of UNC-53 is a component by using a vertebrate homologue of UNC-53 protein of C. elegans. According to this aspect of the invention the method comprises:

(a) contacting an extract of the cell with the vertebrate homologue of UNC-53 protein of C. elegans or a functional equivalent, fragment or bioprecursor of said homologue,

(b) identifying the vertebrate homologue of UNC-53 protein/protein complex formed and

(c) analysing the complex to identify any protein bound to the vertebrate homologue of UNC-53 protein other than the same vertebrate homologue of UNC-53 protein.

This method can also advantageously be used to identify further proteins in a signal transduction pathway of a cell by contacting an extract of the cell used as described above, with any protein identified from step (c) above not being a vertebrate homologue of UNC-53 protein and repeating steps (b) and (c).

Other methods which may be used for identifying proteins in a signal transduction pathway of a cell may comprise for example a western blot overlay method which method is well known to those skilled in the art. Cell extracts are run on gels to separate out protein and subsequently blotted onto a nylon membrane. These membranes may then be incubated, for example in a medium containing vertebrate homologue of UNC-53 having a label attached thereto such as a

biotin or radiolabel and any protein conjugates visualised with for example a streptavidin or alkaline phosphatase conjugated antibody.

The present invention also advantageously provides a process for the preparation of binding antibodies which recognise proteins or fragments thereof involved in the rate and direction of cell migration or the control of cell growth or shape, for the above methods.

The monoclonal antibody for binding to the appropriate vertebrate homologue of UNC-53 (or its functional equivalent) may be prepared by known techniques as described by Kohler R. and Milstein C., (1975) Nature 256, 495 to 497.

Another method which may be used to identify proteins involved in the signal transduction pathway of a cell of which a vertebrate homologue of an UNC-53 protein of C. elegans according to the invention or is a component, involves investigating protein-protein interactions using the two-hybrid vector method. This method, which is well known to those skilled in the art was first developed in yeast by Chien *et al* (1991). This technique is based on functional reconstruction in vivo of a transcription factor which activates a reporter gene. More particularly the technique comprises providing an appropriate host cell with a DNA construct comprising a reporter gene under the control of a promoter regulated by a transcription factor having a DNA binding domain and an activating domain, expressing in the host cell a first hybrid DNA sequence encoding a first fusion of a fragment or all of a nucleic acid sequence according to the invention and either said DNA binding domain or said activating domain of the transcription factor, expressing in the host at least one second hybrid DNA sequence, such as a library or the like, encoding putative binding proteins to be investigated together with the DNA

binding or activating domain of the transcription factor which is not incorporated in the first fusion; detecting any binding of the proteins to be investigated with a protein according to the invention by detecting for the presence of any reporter gene product in the host cell; optionally isolating second hybrid DNA sequences encoding the binding protein.

An example of such a technique utilises the GAL4 protein in yeast. GAL4 is a transcriptional activator of galactose metabolism in yeast and has a separate domain for binding to activators upstream of the galactose metabolising genes as well as a protein binding domain. Nucleotide vectors may be constructed, one of which comprises the nucleotide residues encoding the DNA binding domain of GAL4. These binding domain residues may be fused to a known protein encoding sequence, such as for example a sequence coding for the vertebrate homologue of UNC-53. The other vector comprises the residues encoding the protein binding domain of GAL4. These residues are fused to residues encoding a test protein, preferably from the signal transduction pathway of the vertebrate in question. Any interaction between the vertebrate homologue of UNC-53 protein and the protein to be tested leads to transcriptional activation of a reporter molecule in a GAL-4 transcription deficient yeast cell into which the vectors have been transformed. Preferably, a reporter molecule such as ß-galactosidase is activated upon restoration of transcription of the yeast galactose metabolism genes. This method enables any interactions between proteins involved in the signal transduction pathway or a parallel or redundant pathway to be investigated.

Any proteins identified in the signal transduction pathway of the cell, which may be for example a mammalian cell, may also be included in a

pharmaceutical composition together with a pharmaceutically acceptable carrier, diluent or excipient therefor.

The present invention also provides a process for producing a vertebrate homologue of an UNC-53 protein of C. elegans according to the invention which process comprises culturing the cells transformed or transfected with a cDNA expression vector having any of the cDNA sequences according to the invention as described above, and recovering the expressed protein homologue. The cell may advantageously be a bacterial, animal, insect or plant cell.

A particularly preferred process for producing said vertebrate homologue of UNC-53 protein uses insect cells. Accordingly, the invention provides a process for producing a vertebrate homologue of UNC-53 protein of C. elegans according to the invention which process comprises culturing an insect cell transformed or transfected with a recombinant Baculovirus vector, said vector comprising a nucleotide sequence encoding said vertebrate homologue of UNC-53 protein according to the invention downstream of the Baculovirus polyhedrin promoter and recovering the expressed protein. Advantageously, this method produces large amounts of protein for recovery. The insect cell may be from for example Spodoptera frugiperda or Drosophila Melanogester.

In accordance with the present invention, a defined nucleic acid sequence includes not only the identical nucleic acid but also any minor base variations from the natural nucleic acid sequence including in particular, substitutions in bases which result in a synonymous codon (a different codon specifying the same amino acid), due to the degenerate code in conservative amino acid substitution. The term "nucleic acid sequence" also includes the complimentary sequence to any single stranded sequence

given which includes the definition above regarding base variations.

Furthermore, a defined protein, polypeptide or amino acid sequence according to the invention, includes not only the identical amino acid sequence but also minor amino acid variations from the natural amino acid sequence including conservative amino acid replacements (a replacement by an amino acid that is related in its side chains). Also included are amino acid sequences which vary from the natural amino acid but result in a polypeptide which is immunologically identical or similar to the polypeptide encoded by the naturally occurring sequence. Such polypeptides may be encoded by a corresponding nucleic acid sequence.

A further aspect of the invention provides a nucleic acid sequence of at least 15 nucleotides of a nucleic acid according to the invention and preferably from 15 to 50 nucleotides.

These sequences may, advantageously be used as probes or primers to initiate replication or the like. Such nucleic acid sequences may be produced according to techniques well known in the art, such as by recombinant or synthetic means. They may also be used in diagnostic kits or the like for detecting for the presence of a nucleic acid according to the invention. These test generally comprise contacting the probe with a sample under hybridising conditions and detecting for the presence of any duplex formation between the probe and any nucleic acid in the sample. Nucleic acid sequences according to the invention may also be produced using recombinant or synthetic means such as described in Sambrook et al (Molecular Cloning: A Laboratory Manual, 1989). Advantageously, human allelic variants or polymorphisms of the DNA according to the invention may be identified by, for example, probing DNA from a range of individuals for example from different populations. Furthermore,

nucleic acids and probes according to the invention may be used to sequence genomic DNA from patients using techniques well known in the art, such as the Sanger Dideoxy chain termination method, which may advantageously ascertain any predisposition of a patient to certain disorders.

A method of detecting whether a compound is an inhibitor or an enhancer or expression of a vertebrate homologue of UNC-53 of C. elegans, according to the invention is also provided which method comprises contacting a cell expressing said homologue with said compound and monitoring for a phenotypic change compared to a control cell which has not been contacted with said compound.

Preferably the cell is a transgenic cell as described above. Alternatively the cell may have undergone loss of contact inhibition.

The present method also provides for determining whether said compound is an inhibitor or expression of said vertebrate homologue. In one embodiment the compound to be tested comprises a nucleic acid.

Preferably said nucleic acid sequence comprises an antisense DNA sequence or a mRNA sequence.

Preferably said mRNA sequence comprises 3' untranslated regions of mRNA encoding for said vertebrate homologue.

Alternatively, the compound to be tested may be a protein. Preferably, said protein comprises a protein having an amino acid sequence potentially suitable for inhibiting function of said vertebrate homologue and preferably comprises a protein identified by the methods as described herein.

The present invention also provides a pharmaceutical composition comprising a compound, for example an antisense nucleic acid identified according to the above described method together with a pharmaceutically acceptable carrier, diluent or

excipient therefor.

A nucleic acid sequence or protein identified according to this aspect of the invention may be used as a medicament, or in the preparation of a medicament, for treating loss of contact inhibition of cancer which is mediated by vertebrate homologue of UNC-53 protein or a functional equivalent, fragment, derivative or bioprecursor of said homologue.

Further provided by the invention is a nucleic acid as defined above for use in preparation of a medicament for inhibiting expression of a gene coding for a vertebrate homologue of UNC-53 protein of C. elegans.

Further provided by the invention is an assay for detecting expression of the vertebrate homologue of UNC-53 protein of C. elegans in a vertebrate cell which assay comprises contacting a cell or an extract thereof with an antibody to said vertebrate homologue, which antibody is fused to a reporter molecule, removing any unbound antibody and monitoring for the presence of said reporter molecule.

Preferably the reporter molecule is an antibody conjugated to for example a fluorophore such as fluorescein or alternatively to an enzyme such as strepavidin.

There is also provided a method for detecting for expression of a gene coding for the vertebrate homologue of UNC-53 protein of the invention which method comprises contacting a probe specific for a nucleic acid of protein sequence coding for or corresponding to said vertebrate homologue according to the invention with a cell extract, which probe is linked to a reporter and analysing for the presence of said reporter.

Preferably the probe is a complementary sequence to a region of mRNA transcribed from said gene encoding said vertebrate homologue of UNC-53 protein

according to the invention.

Preferably the complimentary sequence is a 3′ or 5′ untranslated region of said mRNA. Preferably said reporter may be a dig label, a fluorophore, a hapten or a radiolabel.

Alternatively said probe may comprise an antibody specific for said vertebrate homologue of said UNC-53 protein.

Preferably the reporter is an antibody conjugated to for example a fluorophore such as fluorescein or alternatively an enzyme such as streptavidin.

As described above, UNC-53 protein of C.elegans has been found to localise to microtubule and particularly to microtubule (+) ends. Therefore, there is provided by a further aspect of the present invention a method of determining whether a compound is an inhibitor or an enhancer of association of the UNC-53 homologue of the invention to microtubules or plus end regions thereof, which method comprises (a) contacting said compound with a transgenic cell, tissue or organism expressing said vertebrate homologue and which protein is operably linked to a reporter molecule (b) screening for the localisation of said reporter molecule as compared to a cell according to step (a) which has not been contacted with said compound.

A compound identifiable by the above method also forms part of the present invention. Such a compound identified as an inhibitor of localisation or association of said vertebrate homologue with microtubules or the plus end region thereof may be used in alleviating the spread of disease inducing cells or metastasis or loss of contact inhibition. Further a compound identified as an enhancer of association of said vertebrate homologue with microtubules or the plus end region thereof may be used in for example promoting neuronal regeneration,

revascularisation or wound healing, or for treating chronic neurodegenerative diseases or acute traumatic injuries or fibrotic disease. These compounds may then be included in a pharmaceutical composition, together with a pharmaceutically acceptable carrier, diluent or excipient therefor.

Also provided by the present invention is a kit for determining whether a compound is an inhibitor or an enhancer of association of the vertebrate homologue thereof according to the invention with microtubules or the plus end regions thereof, which kit comprises at least one transgenic cell expressing said UNC-53 vertebrate protein homologue and a reporter molecule or a host or transgenic cell according to the invention and at least one cell of the same cell type for use as a control and means for contacting said compound with one of said at least one transgenic cells. Compounds identified as inhibitors or enhancers or microtubule association described above may advantageously be included in a composition and linked to said vertebrate homologue according to the invention to target the compounds to the microtubules or the plus end regions thereof. Such a composition may also comprise, for example, a suitable transfecting or transformation agent.

According to a further aspect of the invention there is provided a method of targeting a protein to a cell microtubule or the plus end region thereof, which method comprises introducing into a host cell, tissue or organism a transgene comprising a sequence capable of expressing said UNC-53 vertebrate homologue according to the invention, which sequence is operably linked to a sequence encoding said protein to be targeted such that a chimeric protein is expressed and which results in targeting of said protein to said microtubule or a plus end region thereof. An even further aspect of the invention comprises a method of

identifying a molecule which covalently modifies UNC-said vertebrate homologue according to the invention, which method comprises a) contacting either an extract from a cell or cells expressing said vertebrate homologue or a mixture of enzymes comprising candidate UNC-53 modifying enzymes in the presence of an indicator of covalent modification of a protein, b) identifying any covalently modified UNC-53 protein from step a) and c) identifying said molecule involved in said modification step. Such an indicator may be ^{32}P.

Further provided by the invention is a method of identifying a compound which alleviates or enhances the toxicity of said UNC-53 vertebrate homologue thereof according to the invention, or which alleviates or enhances apoptosis. The method of the former comprises contacting said compound with a transgenic cell, tissue or organism according to the invention and monitoring for the presence of said reporter molecule adjacent said microtubules or the plus end region thereof. In the case of apoptosis the method comprises monitoring the effect of the compound on cell death.

The invention may be more clearly understood from the following examples which are purely exemplary, with reference to the accompanying drawings wherein,

Figure 1(a) is an illustration of the nucleotide sequence encoding the first human homologue of UNC-53 designated Hs-UNC-53/1 and further variants thereof.

Figure 1(b) is an illustration of the amino acid sequence of hs-UNC-53/1 encoded by the sequences in Figure 1(a).

Figure 1(c) is an illustration of the nucleotide sequence encoding the second human homologue of UNC-53 protein of C. elegans designated Hs-UNC-53/2 and further variants thereof.

Figure 1(d) is an illustration of the amino acid

sequences of Hs-UNC-53/2 encoded by the sequences in Figure 1(c).

Figure 1(e) is an illustration of a nucleotide sequence encoding the third human homologue of UNC-53 protein according to the invention designated Hs-UNC-53/3, and variants thereof.

Figure 1(f) is an illustration of the amino acid sequences of the Hs-UNC-53/3 encoded by the sequences of Figure 1(e).

Figure 1(g) is an illustration of the nucleotide sequence of a genomic DNA fragment that contains a putative 5′ exon of Hs-unc-53/1.

Figure 1(h) is an illustration of the nucleotide sequence AB023155 encoding the protein KIAA0938, a transcript comprising the 3′ half of Hs-unc-53/3.

Figure 1(i) is an overview of the C. elegans and human UNC-53 proteins as cloned. The 5′ truncated variants and a number of the known splice variants have been indicated.

Figure 2 is an alignment of the amino acid sequences of Ce-UNC-53, *Hs-UNC-53/1*, *Hs-UNC-53/2* and *Hs-UNC-53/3*.

Figure 3 is an alignment of the C. elegans unc-53 and the predicted amino acid sequence of C. briggsiae unc-53.

Figure 4 is a list of ProSite signatures for vertebrate UNC-53s based on the sequence alignment.

Figure 5a is an illustration of expression of the three human UNC-53s as studied by Northern blotting.

Figure 5(b) is an illustration of differential expression of Hs-unc-53/3 in different brain parts.

Figure 6(a) is an illustration of differential splice variant expression of Hs-unc-53/1 using RT-PCR.

Figure 6(b) is an illustration of differential splice expression of Hs-unc-53/2 using RT-PCR.

Figure 6(c) is an illustration of differential expression of Hs-unc-53/3 using RT-PCR.

Figure 6(d) is a sequence confirmation of AB023155 expression in cells other than brain using RT-PCR.

Figure 7(a) is an illustration of the cloning of Hs-unc-53/3.

Figure 7(b) is a plasmid map and the nucleotide sequence of the pGI3303 expression vector (C-terminal Hs-unc-53/3 fragment in fusion with GFP).

Figure 7(c) is an illustration of the amino acid sequence of GFP: C-terminal Hs-unc-53/3 fragment (insert of pGI3303).

Figure 7(d) is a plasmid map and the nucleotide sequence of the pGI3305 expression vector (full length Hs-unc-53/3 in fusion with GFP).

Figure 7(e) is an illustration of the amino acid sequence of GFP : Hs-unc-53/3 (insert of pGI3305).

Figue 8 is an illustration of the filipodia and lamellipodia outgrowth of N4 mouse neuroblastoma cells transfected with pGI3303 (F-actin cytoskeleton reorganisation)

Figure 9 is an illustration of the co-localisation of the GFP:Hs-unc-53/3 fusion protein with microtubules in N4 mouse neuroblastoma cells transfected with pGI3305.

Figure 11a is an illustration of the homology domains between Hs-unc-53/3 and a gene encoded (partially) by the Drosophilia melanogaster BAC clone BACR48M05 (AC005719). Results of a TBLASTN search on the non-redundant database with Hs-unc-53/3 as query.

Figure 11b is an illustration of an ORF encoded by the Drosophila melanogaster BAC clone BACR48M05 (AC005719) as predicted by the computer program Fgene.

Figure 11c is an illustration of a "BLAST 2 sequences" search result with Hs-unc-53/3 as query and the Fgene predicted UNC53 homology ORF of D. melanogaster BAC clone BACR48M05.

Figure 12 is an illustration of a zebra fish EST

encoding Dr-unc-53/2.

Figure 13 Genemap98 results for Hs-unc-53/2.

Figure 14 is a schematical drawing of the sequence of the exon containing the putative alternative start codon of human Hs-unc-53/1.

Figure 15 is an illustration of the nucleotide sequence of pGI3150 and the amino acid sequence of the eGFP fusion with a C-terminal fragment of Hs-Unc-53/1.

Figure 16 is an alignment of EST clone yk480b6 and Ce-unc-53 demonstrating a novel splice variant of Ce-unc-53.

Figure 17 is a graphical display of the effect of Hs-unc-53/3 GFP chimera transient transfection on the form factor of N4 cells.

DEPOSITED MATERIAL

Plasmids pG13303 and pG13305 were deposited under accession numbers LMBP3936 and LMBP3937 respectively on 28 May 1999 at the Belgian Coordinated Collections of Microorganisms (BCCM) at Laboratorium voor Moleculaire Biologie - Plasmidencollective (LMBP) B-9000 Ghent, Belgium, in accordance with the provisions of the Budapest Treaty of April 28 1977.

Hs-UNC-53/3 is a bona fide UNC-53 (fig. 1; 2; 3)

Blastn and Tblastn EST-database mining using the sequence of the already known animal UNC-53s led to the identification of 3 ESTs suggestive of novel unc-53s (see experimental procedures). By 3'- and 5'-RACE extension using suitable libraries, it was shown that these ESTs identified a novel unc-53 designated Hs-unc-53/3 (Fig. 1 e; f). The publication of the sequence AB023155 (Nagase et al. 1999, DNA Res. 6:63-70) independently confirmed the correctness of the 3'-end of Hs-unc-53/3 as well as the existence of one new

intron that forms the 5'-end of AB023155. Alignments of the *C. elegans* and 3 human UNC-53 sequences (fig. 2) clearly illustrates that the third human homologue of C. elegans UNC-53 protein is a bona fide UNC-53 with highest similarity to Hs-UNC-53/2 and in decreasing order to Hs-UNC-53/1 and (C. elegans UNC-53) Ce-UNC-53.

Many of the domains of Hs-UNC-53/3 show highest similarity to functional domains of other animal UNC-53s (fig. 2). This critically suggests that Hu-UNC-53/3 most likely has the key functionalities observed for Ce-UNC-53 in a variety of assays including F-actin binding, F-actin reorganisation in cell culture, microtubule and microtubule (+)-end binding in cultured cells, binding of SH3-domain adapters like SEM-5/GRB-2 or other types of binders of proline rich alpha-helices. These results indicate that like Ce-UNC-53, Hs-UNC-53/1, Hs-UNC-53/2, or Hs-UNC-53/3 can be used in a range of biochemical, cellular and animal assays aimed at discovering tissue- or disease-specific modulators of Hs-unc-53 functioning in diagnostic assays.

Further extension of the Unc-53 family (Fig. 11, 12)

Database searches with the three human UNC-53 protein sequences revealed several expressed sequence tags (ESTs) and genomic DNA sequences (BACs) that show significant similarlity to human UNC-53.

C. briggsiae

The C. elegans genome consortium sequenced the locus of the C. briggsiae unc-53 homologous gene. Through gene prediction programs and the cDNA sequence of the C. elegans unc-53, prediction

can be made for the C. briggsiae protein sequence. Alignment of the derived C. briggsiae amino acid sequence with the C. elegans amino acid sequence in figure 3 demonstrates the strong homology of both proteins.

D. melanogaster

BAC clone BACR48M05 (AC005719) clearly contains 3 different exons with high homology to Hs-unc-53/3 (Figure 11). Using the gene structure prediction program Fgene [Solovyev et al., 1995, in: Proceedings of the Third International Conference on Intelligent Systems for Molecular Biology (eds. Rawling et al., Cambridge, England, AAAI Press); Solovyev and Lawrence, 1993, in: Abstracts of the 4th annual keck symposium. Pittsburgh, 47) it was possible to predict an ORF encoded by BAC clone BACR48M05 that shows homology to Hs-unc-53/3 (Figure 11b). However, every Drosophila cDNA partially or entirely encoded by BAC clone BACR48M05 and which contains one or more sequence blocks as indicated in figure 11a should be considered as a family member of the UNC-53 family. A "BLAST 2 SEQUENCE" search indicates that the sequence situated between the three homology blocks that are indicated in figure 11a is less conserved between human and Drosophila (Figure 11c). The predicted ORF of the Drosophila melanogaster UNC53 gene can be used to identify new members of the family. The zebrafish EST fc21d06 (AI658309) shows an identity of 84% and a homology of 92% to Hs-UNC-53/2. It clearly can be considered as a part of the zebrafish homologue of Hs-UNC-53/2 (Figure 12). Finally, a whole series of human ESTs have been placed in public domain databases. To our knowledge, no one has been able to place these ESTs into contigs that describe a true Hs-unc-53 to a level presented in this specification.

The presently available unc-53 sequences - expressed or genomic - further underscore that the unc-53 gene family is a true animal gene family in helminths, vertebrates and arthropods, three major classes of the animal kingdom.

Refined UNC-53 family description based on alignment (fig. 4).

The alignment of the three human and the C. elegans UNC-53 sequences enables the more refined definition of conserved regions in UNC-53s. In figure 4 there are compiled a number of proSite signatures for either the four animal or the three human UNC-53s.

Differential expression of Hu-UNC-53/3 by Northern blot (fig. 5).

To determine in which cells and tissues the vertebrate UNC-53s play a role, a northern blot analysis has been performed. As indicated in the experimental section, relevant probes were amplified and used to visualise in which normal human tissues and in which cancer cell lines the three human UNC-53s were expressed.

1. A cancer cell line RNA blots probed with Hs-Unc53/1.

A Northern blot of poly-A+RNA from several cancer cell lines (Melanoma G361, Lung Cancer A549, Colorectal Adenocarcinoma SW480, Burkitt Lymphoma DRajii, Leukemia Molt4, Lymphoblastic Leukemia K562, HeLa S3 and Promyelocytic Leukemia HL60) was probed using the whole insert of pHH3b. No or weak expression was detected in the Burkitt Lymphoma DRajii, the Leukemia Molt4 and the Promyelocytic Leukemia HL60 cell lines. Five different transcripts

are detected in the remaining cancer cell lines: transcripts 1 and 2 are larger than 9.5kb, transcripts 3 and 4 are 6 to 7 kb and the fifth transcript is around 6 kb. Transcripts 1 and 2 are present in all expressing cell lines but at different levels. Transcripts 3 and 4 are restricted to Melanoma G361, Lung Cancer A549 (weak) and Colorectal Adenocarcinoma SW480 and are the predominant transcripts in Melanoma G361 and Colorectal Adenocarcinoma SW480. Transcript 5 is restricted to Lymphoblastic Leukemia K562 (weak) and (predominant) in HeLa S3 and is predominant in HeLa S3.

2. Cancer cell lines RNA blots probed with Hs-Unc53/2.

A similar set of cancer cell line Northern blots were probed with a 652bp fragment of EST46037 amplified by using the primers 5'-aggagatgaagctgacagatatcc and 5'-aaacaccagtgagtcc. Hs-Unc53/2 is expressed in Melanoma G361, Colorectal Adenocarcinoma SW480, Lymphoblastic Leukemia K562 and HeLa S3. No expression was detected in Lung Cancer A549, Burkitt Lymphoma DRajii, Leukemia Molt4 and promyelocytic leukemia HL60. Interestingly only 2 transcript sizes were detected of around 7 kb expressed in Lymphoblastic Leukemia K562 and HeLa S3 and a transcript of >9.5 kb in Melanoma G361 and Colorectal Adenocarcinoma SW480 and weakly in HeLa53. Noteworthy is the very high expression in melanoma G361.

3. Normal Human tissue probed with Hs-Unc53/1.

A Northern blot of poly-A+RNA from normal human tissue was probed using the whole insert of phage HH3b. Expression levels are low in all tissues with the highest level in heart and placenta, several fold lower levels in brain and testis, even lower

levels in skeletal muscle, pancreas, thymus, colon, small intestine, ovary and prostate. Expression in peripheral blood leukocyte, lung, liver, kidney, spleen is barely detectable.

4. Normal Human tissue probed with Hs-UNC53/2.

A similar set of blots were probed with a 652bp fragment of EST46037 amplified by using the primers 5'aggagatgaagctgacagatatcc and 5'-aaacaccagtgagtcc. Expression levels are low in all tissues with the highest level in kidney, placenta and pancreas, lower levels in heart and lung. Expression is barely detectable or undetectable in skeletal muscle, spleen, thymus, prostate, testis, ovary, small intestine, colon peripheral blood leucocyte, stomach, thyroid, spinal cord, trachea, adrenal gland and bone marrow. Also Hs-unc-53/2 appears to be expressed as different transcripts (figure 5a).

The hs-UNC53/1 and hs-UNC-53/2 homologues are clearly highly regulated genes, showing a strong tissue specificity and, probably, additional mechanisms of regulation (ie differential splicing of different promoters). The different proteins derived from RNA's identified by probe hh15 presumably share the carboxyterminal nucleotide binding domain. Ce-UNC-53 was shown to be a complex genetic locus and complex transcription unit. The different transcripts are thought to be a mechanism to assure the necessary specificity and functional diversity of this signal transduction pathway, with respect to different signals and receptors, different tissues and different directions of migration. The occurrence of a new transcript or the observed changes in expression levels in the cancer cell line blot suggests a role for hs-UNC-53/3 in the establishment or maintenance of the transformed state of those cells.

Expression pattern of hs-UNC-53/3.

A northern blot of poly-A+RNA from several cancer lines was probed with unique fragments of the three genes from the Hs-unc-53 family. Hs-unc-53/3 has a high expression level in lung carcinoma line A549, where only a moderate expression of hs-unc-53/1 has been detected. Furthermore, moderate expression of Hs-unc-53/3 was also observed in melanoma line G361, where previously, a high expression of hs-UNC-53/1 and hs-UNC-52/2 has been observed. This indicated the involvement of hs-unc53/3 in at least two cancer lines.

In normal human tissues, the expression of hs-unc-53/3 shows a clearly new and previously unobserved expression pattern. This difference of expression of hs-unc-53/3 in relation to its homologues hs-unc53/1 and hs-unc53/2 is important for the allocation of functionality to hs-unc-53/3.

Hs-unc-53/3 is highly expressed in brain, as shown on the Northern blots (figure 5a). In figure 5b it can be seen that Hs-unc-53/3 also is differentially expressed in different parts of the brain. Its homologues are not or weakly expressed in brain. This gives an indication that its function in directionality of cell migration and growth cone steering will be in relation to specific regions or cells of the brain. It is deduced that Hs-unc-53/3 will be an important signal transducer or signal adapter linking signals to neuronal outgrowth, axon guidance, and formation and maintenance of synaptic connections. It seems that the function of Hs-unc-53/3 will be associated with neuron-neuron interactions, neuronal outgrowth, neuron muscle interactions, and post-synaptic signal transduction. Furthermore, Hs-unc-53/3 may be involved in the development of cancer of neuronal origin, like

neuroblastomas, or the development of tumours will have their developmental origin in the brain as some eyes diseases like retinoblastomas.

The significance of the high expression of Hs-unc-53/3 in brain tissue can be associated with the high levels of expression which has also been observed in the spinal cord, containing neuronal tissue. Here, neuronal (axon) outgrowth and neuron-neuron connections are of importance. Development of pharmacological tools acting on this pathway may lead to treatments of diseases involved in the growth and movement of neuronal cells, and the regeneration of neuronal connectivity after trauma, or the inhibition of neuronal cancers such as neuroblastomas. Due to its specific expression, inhibitors and/or enhancers specific for Hs-unc-53/3 will have an advantage as a pharmaceutical compound over more general compounds acting on the Hs-unc-53 family of genes and proteins.

A second tissue where hs-UNC-53/3 is highly expressed and where (its) other human homologues are not expressed is the spleen. Hs-UNC-53/3 could therefore function as part of the signal transductions pathway involved in the maturation of leukocytes. Malfunction of this pathway may lead to incorrect maturation of the leukocytes and the development of autoimmune diseases such as rheumatoid arthritis and sclerosis. Next to the signalling function in the recognition of the leukocytes, Hs-UNC-53/3 may also play an important role in the induction and/or signalling pathway of the mechanism underlying apoptosis of leukocytes in the spleen. Pharmaceutical methods involving the hs-UNC-53/3 pathway, which may, for example, result in an inhibition and/or enhancement of its expression may lead to treatment of these disorders. Furthermore, hs-UNC-52/2 may have an advantage, as an inhibitor or enhancer specific for hu-unc53/3 which will act in a more specific manner.

The Hu-UNC-53/3 protein is also highly expressed in the ovary, where the two other human homologues are also expressed. Finally moderate to low expression of hs-unc53/3 is observed in heart, placenta, testis, stomach and adrenal gland.

Although the predominant transcripts of Hs-unc-53/3 are > 9 kb, often a smear occurs that ends at with somewhat higher intensity at 5.5 - 6.5 kB. This short transcript may correspond to AB023155.

The Hs-unc53/3 gene is a highly regulated gene, showing strong tissue specificity and additional mechanisms of regulation which have not previously been identified in any of its known homologues. These findings may thus lead to the development of more specific inhibitors or enhancers of hs-UNC-35/3 and or of the Hs-UNC-53/3 pathway. The Northern blot studies indicate that the three human unc-53s are complex transcriptional units with highly regulated tissue specificity and that transcripts of different lengths exist.

Splice variants of human unc-53s

Whilst cloning Hs-unc-53/3, it became apparent that at least three expression variants of Hs-unc-53/3 - most probably alternative splices - exist (fig. 1e, f; lowercase regions). Targeted efforts for the two other human UNC-53s demonstrated that the other human UNC-53s contained variants (fig. 1a, c and e regions).

Splice variants as observed to date appear to be concentrated in specific regions. A first one (starting at position 1252 in fig. 2) - in which the overall amino acid similarity is weak - contains 2 (splice) variants of both Ce-unc-53 and Hs-unc-53/3. In the worm, the presence or absence of these 2 exons in unc-53 regulates the function of the UNC-53 protein in such a way that cells differentially translate

extra-cellular signal gradient as an attractive or repulsive signal. The most 3'-variant of Hs-unc-53/2 roughly covers the 2 Ce-unc-53 variants.

The complexity of variation in this zone of Hu-UNC-53 might resemble the situation in the nematode. In Hs-unc-53/3, for example, the region from position 3795 to 4325 (figure 1e) consists of two adjacent blocks (3795 to 4283 and 4286 to 4325 in figure 1e) that can independently be present in or absent from cDNAs from frontal cortex tissue. In contrast, no variants were as yet observed in this zone for Hu-UNC-53/1 or /2.

The second variant in Hs-unc-53/3 (fig. 2) deletes a box (MQLDNRTLPKKGLR), which is extremely conserved (in bold) among all human unc-53s. This occurrence of this variant could indicate differentially active functional variants of Hu-unc53/3.

A second region in which splice variants were observed contains a major highly conserved domain of unc-53s. Hs-unc-53/1 has a first variant that comprises the most N-terminal portion of this conserved domain (SGSFRD). A second splice variant in Hs-unc-53/1 (AEERMOSE) lies within the highly conserved domain. Another conserved spot for splice variation in human unc-53s has been found (figure 2): Hs-unc-53/1 {VYE}; -/2 {VNE} and -/3 {NSRGSEL}. All these spliced exons are flanked by two conserved charged domains - putative nuclear localisation signals. Given this conservation, we searched for splice variation in C. elegans and found it to exist in the form of an extra exon (ALSVDSQ) (figure 2). Hu-unc-53/3 has another variant (SPLVWPPKKRQNGPVIYKHSR) (fig. 2).

The most 3' splice variant in Hs-unc-53/3 has been discovered whilst cloning Hs-unc-53/3 and was shown to be present uniquely in human heart cDNA

libraries.

Single nucleotide polymorphisms

Cloning and PCR studies indicated the existence of a non-silent single nucleotide polymorphism in Hs-unc-53/1 in position 1232 and in Hs-unc-53/2 in position 929. This indicated that variations exist in human unc-53s which - in some cases - may be relevant to the proper functioning of the UNC-53 protein and hence in disease.

Expression in normal and neoplastic cells by RT-PCR

The cloning efforts demonstrated the existence of splice variants in the human unc-53s and the Northern blots revealed a range of transcripts for each human unc-53. The combined data do not explain completely the range of transcripts observed. Therefore, our understanding of the expression complexity of human unc-53s may be incomplete and more detailed RT-PCR studies were performed.

One of the obscuring factors could have been that all studies performed on mRNA or cDNA of whole tissues which are built of different normal human cell types that occur in different proportions. For this reason and because skin was not covered in the Northern blot studies, a RT-PCR study was set up using cDNA preparations of the different cells in skin normal human: (1) epidermal keratinocytes, (2) melanocytes, (3) dermal fibroblasts. In addition, lineage matched transformed cell lines or tumour cell lines were included in the study to compare normal versus neoplastic cells. Human umbilical vein endothelial cells (HUVEC) were taken as a normal human match for endothelial cell lines.

The RT-PCR study for Hs-unc-53/1 revealed that the most 5'-splice variant is differentially expressed in normal versus neoplastic cells/cell lines. This exon is present in 7/7 keratinocytes, HUVEC and in melanocytes but lacking in HaCat, ECV304, 2/7 melanoma and MCF-7 cells (breast carcinoma).

The RT-PCR study for Hs-unc-53/2 revealed a more surprising picture. The tumourigenic endothelial line ECV304 lacks expression of Hs-unc-53/2, whereas their normal counterpart HUVEC expresses Hs-unc-53/2, suggesting gene deletion or inactivation of expression in ECV304. In epidermal keratinocytes and the lineage matched spontaneously transformed keratinocyte HaCaT and MCF-7 lack expression of the 5'-end of Hs-unc-53/2, but express the 3'end (starting in or near the microtubule-binding domain). This suggests that like AB023155 for Hs-unc-53/3, also Hs-unc-53/2 can be expressed as a truncated 3'-variant in a cell-specific way. Also splice variation of Hs-unc-53/2 appears to differ in a normal to neoplastic way: the {VNE} exon was shown to be present in all keratinocyte isolates but not in HaCaT and also melanocytes express it, but not 2/7 melanoma or MCF-7. The RT-PCR studies for Hs-unc-53/3 were focussed on demonstrating expression of AB023155 in tissues other than brain. The new exon described was shown to be present in keratinocytes, HUVEC, dermal fibroblasts, melanocytes and their transformed/neoplastic variants, demonstrating its wide expression in tissues in man.

Alternative 5'-start exons

For Hs-unc-53/2 five different start exons have been cloned using RT-PCR, three of which have been confirmed to be present in at least 2 different cDNA libraries (figure 1b, c). Likewise for Hs-unc-53/3 different 5'-exons were found, two of which were

confirmed (figure 1e, f). These 5'-exons most probably indicate that human unc-53s are being expressed via the control of alternative promoters that lie 5' of these different 5'-exons. Also in the nematode has been shown that different (intronic) promoters are driving the expression of 5'-variants of C. elegans unc-53.

The Hs-unc-53/1 5'-end

Despite considerable efforts, cloning has not lead to the identification of a bona fide 5'-end for Hs-unc-53/1 that comprises an F-actin binding domain, despite the fact that the Northern blots indicate the existence of transcripts > 9.5 kb. Given that both Hs-unc-53/2 and -/3 are expressed as full length and truncated forms, the question can be raised whether Hs-unc-53/1 may not be expressed in a short form as well.

cDNA library cloning and 5'-RACE has provided contiguous sequence that ends at a position that matches with a domain in C. elegans un-53, where an alternative start position lies. Based on this argument, Hs-unc-53/1 could be a functional equivalent in man of this transcript in nematode.

To further trace the "longer" variants of Hs-unc-53/1, genomic BAC DNA sequencing has been performed. In figure 1g is shown sequence of a4984 fragment from BAC 585E09. It comprises sequence 5' of the presently known cDNA of Hs-unc-53/1. To the qualified as well as by means of two groups of gene structure prediction computer programs, different but comparable exons in the 4984 bp genomic sequence fragment can be predicted (figure 14). The programs GENSCAN, HEXON and MZEF all predict an exon between bp 1089 and bp 1880. The end of this predicted exon (bp 1880) is confirmed by the cDNA sequence. Therefore this predictions has a big

change to indicate the correct exon length. The programs GRAIL, GENEFINDER and HMMGENE all predict an exon between bp 1123 and bp 2031. None of the predicted exons contains an in frame stop codon 5' of the alternative start codon. Consequently, it is possible that there exist unidentified exons 5' of the exon containing the alternative start codon.

The present picture critically suggests that both nematode and human unc-53s appear to be complex transcriptional units. Moreover, the fact that some of the most complex splice variants map to similar regions in the UNC-53 proteins points to evolutionary conserved functional variants of UNC-53s e.g. with regard to the cells directional migration towards or away from a signal source. In contrast, some of the variants in the human UNC-53s are located in highly conserved domains; these (and other) variants may create discrete - yet undiscovered - functionally different UNC-53 proteins transcribed from one of the unc-53 genes.

The fact that two and maybe three human unc-53s exist as full size and a truncated forms with cell-specific expression, that series of alternative 5'-start exons exist eventually controlled by different promoters that some forms of splice variation are conserved from nematode to man, all indicate that the expression of unc-53s is of very high complexity and that some of the biological functions of UNC-53 proteins are extremely conserved.

On the other hand, the differential expression in Northern blots, the splice variation difference between normal and lineage-matched neoplastic cells and the non-silent single nucleotide changes in two of the three human unc-53s, all indicate how important a wide range of diagnostic assays can be to understand in depth the role in disease of human unc-53s.

Chromosomal localization of Hs-unc-53/2 by Genemap98 (Fig. 13 and 1(c))

The EST clones AA918601, AI248585, AA115014 and AA115015 are clearly homologous to the 3'-UTR of Hs-Unc-53/2 cDNA (Figure 1(c))). Although, AA115014 (describing the same EST as AA115015) contains an alternative splice variant of the Hs-Unc53/2 gene in the 3'UTR. A survey with ESTs AA918601, AI248585, AA115014 or AA115015 as query in the genemap98 database (release November 1998) revealed that the Hs-Unc53/2 gene is located at chromosome 11 (http://www.ncbi.nlm.nih.gov/genemap98/loc.cgi?ID=21224). The STS which is used for chromosomal localization and which is situated in the 3'UTR of the Hs-Unc53/2 gene is referred to as SHGC-33456 (dbSTS Id: 41891, Genbank Acc: G28036, Genbank gi: 1396755) (Figure 13a). The STS was localized by analysis on the NIGMS human/rodent somatic cell hybrid panel (dbSTS Id: 41891). The Radiation hybrid results are summarized in Figure 13b. Together these data imply that every disease or phenotype connected to SHGC-33456 is due to the Hs-Unc-53/2 gene.

Functional Characterisation of Hs-unc-53/3

F-actin reorganisation and microtubule binding of Hs-unc-53/3

Based on its structural features, Hs-unc-53/3 can be classified as a bona fide human unc-53. To further understand its function and in anticipation of developing pharmacological compound screening assays, Hs-unc-53/3 has been physically cloned following the method described in the experimental section and shown in figure 7a. The derived Hs-unc-53/3 clones comprising full length (A to L and the 3'-half (G to

L) of Hs-unc-53/3 were further engineered to form a chimera with green fluorescent protein and cloned into expression vectors appropriate for transfection of eukaryotic cells. The nucleic acid and amino acid sequences of these constructs are shown in figure 7b-e. The constructs were transfected into cells and scored for their effects on the F-actin cytoskeleton and binding to microtubules of mouse neuroblastoma cells N4; functions known for nematode unc-53 and human unc-53/1.

The N4 cell transfected with a GFP fusion to the 3'-half of Hs-unc-53/3 (pGI3303, fig. 7b) showed pronounced filopodia and lamellipodia outgrowth, which is associated with reorganization of the F-actin cytoskeleton (Figure 8). This observation demonstrates that like nematode unc-53 and human unc-53/1, the F-actin binding domain is not required for inducing reorganization of the F-actin cytoskeleton of N4 cells. In addition, the pGI3303 encoded fusion protein does not co-localize with microtubuli but localizes to the cytoplasm of N4 cells indicating that an important domain for microtubuli association is missing in this C-terminal fragment of Hs-unc-53/3. In the alignment figure 2 can be seen that the C-terminal half of Hs-unc-53/3 (approximate KIAA0938) does not comprise the conserved microtubule binding domain.

In contrast, the N4 cells that expressed low to medium levels of the GFP fusion to full length Hs-unc-53/3 (pGI3305, Fig. 7d) displayed a co-localization of the GFP fusion protein with microtubules (Figure 9). Even the centrosomes could clearly be detected in some transfected cells. Cells expressing very low amounts of the fusion protein displayed specific microtubule (+)-end binding (Figure 9). The morphology of the pGI3305 transfected N4 cells does not clearly differ from the control transfected cells although there is a

tendency towards rounding up of the pGI3305 transfected cells and filopodia outgrowth.

Validation of functional assays as compound screens

R74288 has previously been shown to be an inhibitor of nematode function in C. elegans (WO96/38555), an activity that has been confirmed in Ce-unc-53 transfected N4 cells, where only the transgene-induced effect was inhibited by R74288. In order to confirm compound R74288s activity in a full mammalian system, a stable transfection of plasmid pGI3150 was performed in the N4 neuroblastoma cell line with the lipofectamin procedure (Gibco BRL). pGI3150 expresses an eGFP protein in fusion with the C-terminal end of Hs-unc-53/1 (see Figure 15a). After two weeks of G418 selection, 20 clones with stable integration of the pGI3150 plasmid were selected and isolated. These clones were tested for GFP expression by fluorescence microscopy and by Western blotting with an anti-GFP antibody (table 1). The lamellipodia outgrowth phenotype was checked visually (See Figure 15b). Compound R74288 was tested on four random selected pGI3150 stably transfected clones: 8.1, 8.2, 8.3 and 10.1 and on a pool of pEGFPC1 stable transfected N4 control cells. Clones 8.2 and 10.1 displayed less lamellipodia outgrowth than clones 8.1 and 8.3. Compounds and solvents were added to the stably transfected cells (10^{5}M in DMSO). After 24 hrs of incubation, two persons independently scored the effect of the treatments on the cells. As shown in table 1, both persons noticed an effect compound 2 on clones 8.2 and 10.1 with a weak transgene-induced lamellipodia phenotype. This effect consisted of a more flat morphology of the treated versus untreated cells. Compound 2 was R74288.

Table 1. Effect of compounds on lamellipodia formation

Clone	Compound 1	Compound 2	Compound 3	Compound 4	GFP fluo	GFP Western	Phenotype
8.1	0	0	0	toxic	+	+	+
8.2	0	+	0	toxic	++	+++	+/-
8.3	0	0	0	toxic	++	++	++
10.1	0	+	0	toxic	+/-	+	+/-
GFP pool.	0	0	0	toxic			-

Automated compound screening by measuring cell morphology

Compound screening assays must have a sufficiently high throughput to be relevant to drug discovery. To achieve this goal, we automated the procedure of measuring the morphological changes induced in cells following transient transfection with full length or 3'-half of Hs-unc-53/3 GFP chimeras. The cell culture, transfection, fluorescence staining and microscopy procedures are performed within a 96-well plate (all-in-one). The fluorescent staining method comprises a triple fluorescent labeling procedure (1) for cell nucleic using DNA double helix intercalating dyes such as Hoechst 33342 or DAPI, (2) for transfection efficiency and expression level of the chimeric protein using GFP fluorescence and (3) for the F-actin cytoskeleton using fluorescently labeled phalloidin, a microfilament dye.

These three different fluorescent images are collected using an motorised stage plus stage driver and a frame grabber that produces seamless composite images of the cells in the well. The software programs to drive this operation are known in public domain as "SCIL" (University of Amsterdam). The seamless images are then superimposed using

pseudocolour for the operator to inspect the quality of the culture. In addition, the SCIL program was compiled in such a way that it: (1) identifies cells by means of their nucleus, (2) measures the GFP fluorescence intensity, (3) delineates the area of the F-actin (phalloidin)staining surrounding a nucleus and (4) calculates a range of parameters objectively representing the features of the F-actin staining pattern of each individual cell. One example of such a parameter is called the "form factor". It is an arbitrary value that reflects the dendricity of a cell. It is derived by calculating (A) the true circumference of a cell's F-actin staining area as seen in the image and (B) the area of the F-actin staining of that given cell. The ratio $4xPIx(B)^2$ = the form factor. For a rounded cell, the form factor approximates 1 whereas, for a cell with increased filopodia and lamellipodia outgrowth, the true circumference will be much larger than that of a circle and as a result, the form factor << 1.

In experiments it was shown that transiently transfected N4 cell populations indeed displayed a different form factor versus control cells. Both the median and average form factor for a cell population in a well were reduced following transfection with the 3'-half of Hs-unc-53/3. More in particular, there was a significant decrease in the number of cells in a transfected culture that displayed the minimal form factor, suggesting that the Hs-UNC-53/3 transgene induced round cells in particular to become more dendritic (figure 16).

Chromosomal localisation of Hs-unc-53/3 by FISH indicative for a role disease

With FISH technology using a unique fragment of hs-unc-53/3 we are able to localize the hs-unc53/3

gene on chromosome 12q21.1. Chromosome 12q21.1 is a region shown to be involved in autosomal dominant, cornea plana and closed angle glaucoma (Sigler-Villanueva et al., Ophthalmic Genetics 18:55-62, 1997). This indicates that hs-UNC-53/3 protein may be involved in eye development and thus eye diseases, such as retinoblastomas. Neuroblastoma cell line NPG and liposarcoma line WDLPS and other sarcoma lines have amplifications in this region. The neuroblastoma amplification seems to be located more distal (12q24) while the liposarcoma line is located at 12q21 (Van Royal et al., Cancer Genetics and Cytogenetics 82:151-4, 1995). Three loci related to Darier's disease, an autosomal dominant genodermatosis disease characterized by epidermal acantholysis and dyskeratosis have been mapped in region 12q21-q24 (Wright et al., Journal of Investigative Dermatology 103:665-8). 12q21 is also known to be a fragile site associated with the pathogenesis of non-Hodgkin's Lymphoma (Chary-Reddy et al., Cancer Letter 86:111-7 1994). Duplications related to nephroblastoma tumorgensis were commonly found in the 12q21-q23 region (Austruy et al., Genes Chromosomes Cancer 14:285-294, 1995). In a girl with mental retardation, a conclusive disorder and clinical findings resembling cerebral palsy, positioning of segments from other autosomes adject to the band 12q21 were found (Biederman et al., Ann Genet 19:257-260, 1976). Cytogenetic analysis for myeloid leukemia showed a complex caryotype with chromosomal breakpoints at 12q21 (Weinstein et al., Cancer Genet Cytogenet 48:75-81, 1990). Finally, analysis of complex chromosomal rearrangements in malformed children and from spontaneous abortions showed specific breakpoints at site 12q21 Gorski et al., Am J Med Genet 29:247-261, 1997). Most of these diseases have been shown to be involved with cell movement, aberrant development, or

cell-cell contact and neuronal tissue or neuronal development.

Confirmation of FISH with Radiation hybrid panels

To confirm and refine the chromosomal localisation of the human unc-53s an alternative method for FISH has been used. Radiation hybrid (RH) mapping is a somatic cell hybrid technique that was developed to construct high-resolution, contiguous maps of mammalian chromosomes. RH mapping provides a method for ordering DNA markers spanning millions of base pairs of DNA at a resolution to easily obtained by other mapping methods. Some of the advantages of RH mapping are (1) distance estimated by this method is directly proportional to physical distance, (2) nonpolymorphic DNA markers, that can not be used for meiotic mapping, can be used for this method, and (3) a high resolution map that is not easily made by other methods can be obtained.

The results of FISH and RH mapping for the three human unc-53s are summarised in table AA. By using publicly available databases (see experimental section) one can derive information on the correlation between FISH and RH mapping. RH Mapping was shown in this way to confirm the FISH data for the three unc-53s.

Table 2. RH Mapping Primers and Results

Unc-53	FOR Primer	REV primer	PCR Results	Marker*	FISH
Hs-UNC-53/1 (BAC585E9)	5'TGTGGGT GAGGAATGC TGAC	5'CAGAGCTT GCTCTAGAGG AC	51, 62, 66	SHGC-30236	1q31-32
Hs-unc-53/1 (BAC585E9)	5'CCTGCCC AACATAGCA AGAC	5'CCATCTAC AATGAGCCAG AC	51, 62, 66	SHGC-30236	1q31-32
Hs-unc-53/2 G411	5'CTGCCTC CCTTTGCTG TGTTGCATG	5'CTGAGCAG AGTGAAGCCA GAGTTGG	8, 28, 29, 43, 44, 51, 59, 66, 70, 77, 83	AFM022th2	11p15.t
Hs-unc-53/2, F4.1.2	5'TCATGTA TTCCCCACA GACAAGCC	5'CATTGTGT CTTGATACTT TGGGGTGC	8, 28, 44, 51, 59, 65, 83	SHGC-31021	11p15.1
Hs-unc-53/2, D4.1.1	5'GAGGATT TTATTTCTG GGAAATGGA ATCGG	5'TGATCTTC CACTCCGTGG ATAACT	8, 27, 28, 29, 43, 44, 51, 59, 65, 70, 83	AFM022th2	11p15.1
Hs-unc-53/2, J4.1.4	5'AAAGCCC AAGCCCCGG GAGAAGATG	5'AACCCGTT TTCCACCGAG CCGCTC	8, 27, 28, 43, 44, 51, 59, 66, 70, 83	AFM022th2	11p15.1
Hs-unc-53/3, A215	5'ACTTGCT GAAACAGAG AGCTCCATG	5'CTTGCTGT CTTCTTTCTC CTTGGC	1, 48, 50, 51, 59, 65, 66, 73, 74, 76, 78	SHGC-17536	12q21.1
Hs-unc-53/3, A211	5'TGATCTT CTAGCGTGT GACTCACTG	5'ATCATTCC TTGGAGT	1, 48, 50, 51, 59, 73, 76, 78	SHGC-17536	12q21.1

(*) list not exhaustive

Also sequence information available in public domain can help refine the positioning of the unc-53 genes, like in the following example. The EST clones AA918601, AI248585, AA115014 and AA115015 are clearly homologous to Hs-Unc53/2 cDNA. Although, AA115014 (describing the same EST as AA115015) contains an alternative splicevariant of the Hs-Unc53/2 gene in the 3'UTR. A survey with ESTs AA918601, AI248585, AA115014 or AA115015 as query in the genemap98 database (release November 1998) revealed that the Hu-

unc53/2 gene is located at chromosome 11 (http://www.ncbi.nlm.nih.gov/genemap98/loc.cgi?ID=21224). The STS which is used for chromosomal localization and which is situated in the 3'UTR of the Hs-Unc53/2 gene is referred to as SHGC-33456 (dbSTS id: 41891, Genbank Acc: G28036, Genbank gi: 1396755) (Figure 13). The STS was localized by analysis on the NIGMS human/rodent somatic cell hybrid panel (dbSTS id: 41891). The radiation hybrid results are summarized in Figure 13. Together these data imply that diseases or phenotypes connected to SHGC-33456 is due to the Hs-Unc53/2 gene.

EXPERIMENTAL PROCEDURES

Cloning & sequencing of Hs-unc-53/3

Hs-unc53/3 has been cloned starting from a series of ESTs that were similar but not identical to Hs-unc-53/1 or -/2. The ESTs were:

1. WashU-Merck EST 767735.

Transformed cells carrying the EST 767735 sequence were ordered from Research Genetics. Plasmid DNA was isolated using standard protocols (Qiagen plasmid DNA isolation kit), the sequence of the insert was determined.

2. ATCC cDNA clones 86459.

Transformed cells carrying the cDNA clone 86459 sequence were ordered from ATCC. Plasmid DNA was isolated using standard protocols (Qiagen plasmid DNA isolation kit), the sequence of the insert was determined.

3. Genethon cDNA clone c09a03 from the Geneexpress cDNA program.

Transformed cells carrying the cDNA clone c09a03 sequence were ordered from Genethon. Plasmid DNA was isolated using standard protocols (Qiagen plasmid DNA isolation kit), the sequence of the insert was determined.

These ESTs were extended to form one ORF as follows:

1. 5' extension of EST 767735 by RACE (Rapid Amplification of cDNA Ends).

Marathon-Ready cDNAs (Clontech) are premade "libraries" of adaptor-ligated double-stranded cDNA ready for use as templates in RACE experiments. Five ml Marathon-Ready cDNA was used as template in a regular 50 ml RACE. The RACE mixture contained 1 x KlenTaq PCR buffer. 0.2 mM of each dNTP, 1 x advantage KlenTaq polymerase mix (Clontech), 0.15 mM AP1 adaptor primer and 0.15 mM RACE gene specific primer. The amplification conditions were as follows: 94°C for 30 s and 68 °C for 4 min. One-hundred-fold diluted RACE product was used as a template in a nested PCR with AP2 adaptor and gene specific nested PCR primers. Specific nested PCR fragments were cloned into pCR2 (TA cloning kit, Invitrogen) and the sequences of the inserts were determined. Gene-specific primer (hh3UNC53 97101702): 5'ACCATTTACACCTGAAGACGATTGAGGTCC; nested gene-specific primer (hh3UNC53 97101701) 5'CTCCTATTTAAATTAGAGGCTCCCTGGACC Marathon cDNA library: human placenta, human heart, human chronic myelogenous leukemia, human colorectal adenocarcinoma.

2. 3′ extension of EST 767735 by RACE.

Method as described previously. Gene specific primer (hh3UNC53 97102702) 5′CAATCGTCTTCAGGTGTAAATGGTAACGTG; nested gene specific primer (hh3UNC53 97102703) 5′GAATGTCAAACACAGTGCCACCTCCACC Marathon cDNA library: human placenta, human heart, human HeLa, human melanoma.

3. 3′ extension of cDNA clone c09a03 by RACE.

Method as described previously, gene-specific primer (hh3UNC53 98020401) 5′AGGGAGCACTGAATGGTCCAGACCATCCTC; nested gene-specific primer (hh3UNC53 98020402) 5′GCATCAGAAGACAGCATTCCTCTGAAAGTG Marathon cDNA library: human placenta, human heart, human HeLa, human melanoma, human colorectal adenocarcinoma, human chronic myelogenous leukemia.

4. 5′ extension of cDNA clone 86459 by RACE (1).

Method as described previously gene-specific primer (hh3UNC53 98020403) 5′TTCAATTTCTATCTCTATGAGTTTTCTTCG; nested gene-specific primer (hh3UNC53 98020404) 5′GCAGCTCTAGATTTGGTGATGAAGAAACTC Marathon cDNA library: human placenta, human heart, human HeLa, human melanoma. Overlapping sequences were assembled in a single contiguous sequence.

5. 5′ extension cDNA clone 86459 by RACE (2).

Method as described previously gene-specific primer (hh3UNC53 98022502)

5'TCAGAATGTGATGAAGGAGGCTTGGTGGAC; nested gene-specific primer (hh3UNC53 98022501)
5'GGATGCCGGAAGGGATGAATCAGTAAGC Marathon cDNA library: human placenta, human heart, human HeLa, human melanoma, human colorectal adenocarcinoma, human chronic myelogenous leukemia.

Validating variants at 5' end of the cDNA sequence

In the final 5' RACE experiment, 2 variants have been found whose sequence diverge upstream from the IYTDWAN protein sequence (position 289 in figure 1e or position 82 in figure 1f). By using primers ATTTACACTGACTGGGCCAAC and ATAATCTGGATGATTTCTGCTAGGAGT on cDNA clones a Hs-unc-53/3 specific PCR product was obtained that was radiolabeled using the random primed DNA labeling kit (Roche Molecular Biochemicals) and hybridized to human DNA BAC filters (Research Genetics). Both primers are located near the IYTDWAN box. Four BACs turned out positive (415J11; 464C17, 525C02 and 537B02). DNA sequencing of the region upstream from the IYTDWAN protein sequence directly on these BACs showed that this region was preceded by a putative intronic sequence as evidenced by the multiple stop codons in the reading frame and by the consensus AG intron acceptor sequence. For sequencing purposes, BAC DNA was prepared according to a modified Qiagen plasmid DNA procedure.

A primer pair was designed specifically to amplify the 5'end of the variant shown in full in figure 1e (primers ACTTGCTGAAACAGAGAGCTCCATG and CTTGCTGTCTTCTTTCTCCTTGGC). PCR with these primers on BAC DNA showed the presence of the genomic sequence encoding this variant in 3 out of the 4 BACs (not present in BAC 415J11).

BACs containing the genomic sequence encoding the other 5'end variant of Hs-unc-53/3 as shown as the variant in figure 1e were identified by hybridizing the Research Genetics human DNA GAC filters with primer TGATCTTCTAGCGTGTGACTCACTG, radioactively labeled using gamma-P32-ATP and polynucleotide kinase. Positive BACs were 404F14, 450K18 and 764L15.

Sequencing directly on the respective BACs in the 3' direction from within the 2 alternative 5' exons and comparison of the genomic DNA sequence with the previously determined cDNA sequence identified the GT intron donor site. Joining of the genomic sequences from both 5' exons and the IYTDWAN encoding sequence after removal of the predicted intronic sequence restored for both variants the sequence of the 5' RACE experiment without affecting the translation of the Open Reading Frame.

Cloning of Hs-unc-53/3 constructs

With the aim of cloning the full-length Open Reading Frame of Hs-unc-53/3, primer pairs were selected such that the ORF could be amplified in 6 overlapping fragments ranging in size from 1 to 2 kbp. Overlaps between the fragments were chosen such that they contain an endonuclease restriction enzyme recognition site suitable for cloning the full-length gen. For the 5' fragment, the downstream oriented primer was chosen to contain the first putative start codon (ATG) in variant 1 (the one shown in full in figure 1e). PCR conditions using the Expand High Fidelity PCR system (Roche Molecular Biochemicals) for all of the fragments were as follows. Initial denaturation for 5' at 95°C; 30 cycles of denaturation at 95°C for 45", primer annealing at 55°C for 45" and extention at 72°C for 1' (3' for primer combination

A+B); followed by an additional incubation for 7' at 72°C and storage at 4°C. PCRs were run on PE Biosystems 9700 PCR machines.

Primer pairs used for cloning Hs-unc-53/3 fragments

#	Size (bp)	Primer	Sequence
A-B	2229	A	TCAGCTCGAGCATATGCCTGTTCTTGGGGTTGC
		B	GGGGTGGGTCGACTTGTCAAGTGG
C-D	847	C	ATGGAAGGACCATACCCAACTTGAC
		D	CTTGTTCCAGCTTTCTGCCTAGATG
E-F	781	E	CAGGTTCCTGGAGAAGAGGCATGTC
		F	GGTGAGGCAATATCTGGATACTTGG
G-H	1291	G	AGGCAGCCAGGATCCAAGTATCCAG
		H	TGCGAAGATCTTTTGGGAGGATGGTC
I-J	1022	I	AACCATTGAAATGCTGAAGGCTCAG
		J	GGTTATGGGATCTAATTAAGTCTCC
K-L	1255	K	CACTAGCCTTGGTCTGAGCTCTGAC
		L	TCACCCTCTAGAGGGTAGATTCAAG

Primer A contains restriction sites (XhoI and nheI) suitable for final subcloning in an eukaryotic expression vector (pEGFPc3) and in a yeast-two-hybrid vector (pAS2-1), respectively.

PCR products were analyzed by agarose gel electrophoresis and were visualized by ethidium bromide staining. Splice variants as mentioned in figure 1e were observed as multiple bands on agarose gels. Single band PCR products were purified with the Qiaquick PCR purification kit, whereas multiple band PCR products were cut out from gel as individual bands and purified using the Qiaquick gel extraction kit. PCR products were cloned in pCR2.1 according to the suppliers protocol (Invitrogen). For each fragment, multiple clones were picked from selective LB agar plates and grown overnight under antibiotic selection pressure for DNA preparation either on the biorot 9600

(Qiagen), or manually on anion exchange columns (Qiagen tip 20 or tip 100). Insert sequences were determined using the Bigdye terminator ready reaction cycle sequencing kit (PE Biosystems). Individual sequencing reactions for each clone were assembled in single sequence contigs using the Sequencher software package (GeneCodes). Sequences were compared to the previously determined consensus sequence using the SeqEd software package form PE Biosystems. For each fragment a clone was selected containing the correct sequence and the splice variant of interest. For the I-J fragment, a clone was selected that missed the hart specific 22 amino acid splice variant (figure 1f). In the K-L fragment clone, a SfiI-SacII linker was cloned in the BamHI site of the pCR2.1 multiple cloning site to facilitate subcloning of the full-length gene into the yeast-two-hybrid vector (pAS2-1) and the eukaryotic expression vector (pEGFPc3), respectively.

The overall cloning strategy of the full-length gene is visualized in figure 7a. 7a1 illustrates the overlapping PCR fragments and the nomenclature of fragments and primer pairs. 7a2 illustrates the assembly of the 3'half of the gene in pCR2.1. Internal BamHI (I-J fragment) and XhoI (K-L fragment) sites as well as restriction sites from the multiple cloning site of pCR2.1 (as shown in the figure) were removed by side-directed mutagenesis (SDM) using the Quickchange Site-Directed mutagenesis kit (stratagene). The NotI-EcoRI G-H fragment and the EcoRI-NheI I-Jd22 (d22 indicating that the 22 amino acid splice variant is absent) were directionally cloned in the NotI and NheI sites of the K-L fragment clone. Multiple clones were picked and verified by DNA sequencing. 7a3 illustrates the assembly of the 5'half. Internal XhoI (C-D fragment) and SfiI and XhoI (E-F fragment) sites were removed by SDM.

Inserts were cut out from the vectors by restriction digestion with the appropriate restriction enzymes (XhoI+SalI; SalI+NarI and NarI+BamHI, respectively) and purified from gel after agarose gel electrophoresis. The 3 fragments were ligated together, re-cut with XhoI and BamHI and separated on gel. The band of the expected size was cut out of gel, purified and cloned in front of the 3' half, opened by digestion with XhoI and BamHI (figure 7a4). Multiple clones were picked and verified by sequencing.

Figure 7a illustrates the modular nature of the cloning project. For all the possible combinations of splice variation within the building block fragments, one representative clone is available. In view of functional analysis, building blocks can be exchanged easily by standard technology, either in the pCR2.1 construct or in the final eukaryotic expression or yeast-two-hybrid construct.

Construct of Hs-unc-53/3 GFP chimeras

The construction of the mammalian expression vectors pGI3303 and pGI3305 is explained in the legends of figure 7a, 7b and 7d. pG13303 can be used to over-express in mammalian cells or animals a fusion protein between eGFP and 1128 AA C-terminal fragment of Hs-unc-53/3 (Fig 7c). pG3305 can be used to overexpress in mammalian cells or animals a fusion protein between eGFP and the 2363 AA full length Hu-unc-53/3 (fig 7d). The Hs-unc-53/3 cDNA in pGI3303 as well as in pGI3305 contains silent mutations that introduce or remove specific restriction sites in order to be able to easily subclone different types of alternative splice variants in these vectors.

Genomic DNA sequencing (BAC 585E09)

Using the primers AGGACCCTATGCGGAGGTCAAGCCGC and TGGGTTGGCATCATCGCTGTCGTAGC, a PCR specific for Hs-unc-53/1 was developed. PCR products were radiolabeled using the Random Prime DNA labeling kit (Roche Molecular Biochemicals) and hybridized on the human genomic DNA BAC filters (Research Genetics). Positive signals were obtained for BAC clones 366H21, 483L14, 471J09 and 585E09. BAC DNA was isolated from E. coli genomic clone 585E09 according to a modified Qiagen plasmid DNA preparation procedure. A shotgun library of 1920 clones was constructed at GATC (Konstanz, Germany). BAC DNA was prepared, nebulized and subcloned after end-repairing in the sequence vector pTZ19R. At JRF, DNA was prepared on the Biorobot 9600 (Qiagen) from 1440 clones. End sequencing reactions with M13 forward (TGTAAAACGACGGCCAGT) and reverse (CAGGAAACAGCTATGACC) primer were done on 768 clones. 672 additional clones were sequenced with M13 only. 5 μl DNA was used in 15 μl final reaction volume using the BigDye Terminator Ready Reaction sequencing kit. Sequencing reactions were run on MJ Research PTC200 PCR machines. Reaction products were run and analysed on PE ABI 377 DNA sequencers. All sequencing results were imported in the Sequencher (GeneCodes) software package. Contaminating vector sequences and trailing sequences of low quality were trimmed. Individual sequences were assembled in contigs with standard software settings. A great number of contigs were constructed ranging from below 500 bp to over 10 kbp. Singletons are also still present. By looking for strings of known sequence, a contig was found containing the known and reliable 5'end of hUNC53h1 and extending this sequence in 5' direction. This sequence and its relevant features are described in figure 1g and its legend.

Northern blotting

A Human multiple tissue Norther (MTN-1, Clontech) containing in each lane 2 mg of poly A + RNA from eight different human tissues (heart, brain, placenta, lung, liver, skeletal muscle, kidney, and pancreas) and a MTN-II human multiple tissue Northern, containing in each lane 2 mg of poly A + RNA from spleen, thymus, prostate, testis, ovary, small intestine, colon and peripheral leukocyte, were hybridized according to the manufacturer's instructions and washed out in 0.1xSSC:0.2% SDS at 55°C. Also from Clontech, a poly A + RNA blot from human cancer cell lines (melanoma G361, lung carcinoma A549, colorectal adenocarcinoma SW480, Burkitt's lymphoma Raji Leukemia Molt 4, lymphoblastic leukemia K562, HeLa S3 and promyelocytic leukemia HL60) was tested.

Cancer cell lines RNA blots probed with Hs-unc-53/3

A set of cancer cell line Northern blots were probed with a 665 bp fragment of Hs-unc-53/3 amplified by using the primers 5'AGGAATTAAAATTAACGGATATTCGG and 5'AAAACTGTCCAAACTATTTTCTTCTACC. HU-unc-53/3 is expressed in Melanoma G361 and lung carcinoma A549, transcripts sizes were detected of >0.5 kb. No expression was detected in promyelocytic leukemia HL-60 HeLa cell S3, chronic myelogenous leukemia K-562, leukemia MOLT-4, Burkitt's lymphoma Raij and colorectal adenocarcinoma SW480.

Normal human tissue RNA blots probed with Hs-unc-53/3

A set of normal human tissue Northern blots were

probed with a 665 bp fragment of Hs-unc-53/3 amplified by using the primers 5' AGGAATTAAAATTAACGGATATTCGG and 5' AAAACTGTCCAAACTATTTTCTTCTACC. High expression levels were detected in brain, spleen, ovary and spinal cord, lower levels in heart, placenta, testis, stomach, and adrenal gland. Transcripts sizes were >= 9.5 kb.

FISH

Hs-UNC-53/3 is localised to chromosome 12q21.1

Slides preparation:

Lymphocytes isolated from human blood were cultured in α-minimal essential medium (MEM) supplemented with 10% foetal calf serum and phytohaemagglutinin (PHA) at 37°C for 68-72 hr. The lymphocyte cultures were treated with BrdU (0.18mg/ml Sigma) to synchronise the cell population. The synchronised cells were washed three times with serum-free medium to release the block and recultured at 37°C for 6 hr in a α-MEM with thymidine (2.5μg/ml: Sigma). Cells were harvested and slides were made by using standard procedures including hypotonic treatment fix and air-dry.

In situ hybridisation and FISH detection:

A cDNA probe was biotinylated with dATP using the BRL BioNick labelling kit (15°C, 1 hr) Heng et al, 1992). The procedure for FISH detection was performed according to Heng et al., 1992 & Heng and Tsui, 1993. Heng et al..: Proc Natl Acad Sci USA 89: 9509-9513 (1992). Heng et al. Chromosoma 102: 325-332 (1993). Briefly, slides were baked at 55°C for 1 hour. After RNase treatment, the slides were denatured in 70%

formamide in 2xSSC for 2 min. at 70°C followed by dehydrated with ethanol. Probes were denatured at 75°C for 5 min. in a hybridisation mix consisting of 50% formamide and 10% dextran sulphate. Probes were loaded on the denatured chromosomal slides. After over night hybridisation, slides were washed and detected as well as amplified. FISH signals and the DAPI banding pattern were recorded separately by taking photographs, and the assignment of the FISH mapping data with chromosomal bands was achieved by superimposing FISH signals with DAPI banded chromosomes (Heng et al, 1993).

Results

Under the condition used the hybridisation efficiency was approximately 67% for this probe (among 100 checked mitotic figures, 67 of them showed signals on one pair of the chromosomes). Since the DAPI banding was used to identify the specific chromosome, the assignment between signal from probe and the long arm of chromosome 12 was obtained. The detailed position was further determined in the diagram based on the summary from 10 photos.

Radiation Hybrid Mapping

Radiation hybrid analysis is a PCR technique and the panels of radiation hybrid DNA are provided at a concentration of 25 ng/μl in TE buffer suitable for these reactions. Typically, 25 ng of DNA is used in a 10 μl PCR reaction.

Some of the radiation hybrid panels are supported by an e-mail server which can assist you in the chromosome localization of markers. A server for the chromosome localization of markers using the Stanford G3 and Stanford TNG panels is available at http://www-

shgc.stanford.edu. At the time of catalog publication, the Stanford TNG server was capable of chromosome localization only on chromosomes 2, 4, 7 and 21. Chromosome localization of markers from the GeneBridge4 panel may be performed by accessing the server at http://www-genome.wi.mit.edu. RH mapping involves the statistical analysis of several to many markers to determine the relative order of the markers with respect to one another. RH mapping can be achieved using statistical programs that will provide the best map along with a measure of the relative likelihood of one order versus another.

This type of analysis has been shown to successfully generate the order of markers on the RH map that is significantly more likely than any alternative order. Two statistical programs for RH mapping can be downloaded from the World Wide Web free of charge. SAMapper was produced at the Stanford Human Genome Center and be downloaded at http://www-shgc.stanford.edu/Mapping/SAMapper/index.html RHMAP was written by Michael Boehnke at the University of Michigan and can be downloaded at http://www.sph.umich.edu/group/statgen/software. A comprehensive web page regarding radiation hybrid mapping, with links to web sites with analysis software and other information, can be found at http://linkage.rockefeller.edu/tara/rhmap/

Transfection protocol for cells

N$ neuroblastoma lines were seeded in Lab Tek chambered coverglass (Nalgene Nunc International) and transfected with pEGFP (control), pGI3303 and pGI3305 using lipofectamine (Life Technologies BRL). After 24-48 hours, the chambered coverglasses were placed on an inverted fluorescence microscope where GFP fluorescence could be visualized in living cells. The

details of this method have been described in PCT/EP96/02311.

Microscopy and fluorescence staining using phalloidin

have been described earlier (EP97/06956).

SEQUENCE LISTING

Seq ID No 1 is a nucleic acid sequence of Hs unc-53/1 and lacking the nucleotides from position 2873 to 3043 shown in Fig. 1a.

Seq ID No. 2 is a nucleic acid sequence of Hs unc-53/1 and lacking the nucleotides from position 3098 to 3121 shown in Figure 1a.

Seq ID no. 3 is a nucleic acid sequence of Hs-unc-53/1 and lacking the nucleotides from position 3518 to 3526 of the sequence identified in Fig. 1a.

Seq ID No. 4 is an amino acid sequence of Hs-unc-53/1 protein and lacking the amino acids from position 958 to 1014 of the sequence identified in Fig. 1b

Seq ID No. 5 is a amino acid sequence of Hs-unc-53/1 protein and lacking the amino acids from position 1033 to 1040 of the sequence identified in Fig. 1b.

Seq ID No. 6 is a amino acid sequence of Hs-unc-53/1 protein and lacking the amino acids from position 1173 to 1175 of the sequence identified in Fig. 1b.

Seq ID No. 7 is a nucleotide sequence encoding Hs-unc-53/2 and lacking the nucleotides from position 5425 to 5433 of the sequence illustrated in Fig. 1c.

Seq ID No. 8 is a nucleotide sequence encoding Hs-unc-53/2 and lacking the nucleotides from position 5924 to 6024 of the sequence illustrated in Fig. 1c.

Seq ID No. 9 is a nucleotide sequence encoding Hs-unc-53/2 and having the sequence of variant 1 illustrated in Fig. 1c.

Seq ID No. 10 is a nucleotide sequence encoding Hs-unc-53/2 and having the sequence of variant 2 illustrated in Fig. 1c.

Seq ID No. 11 is a nucleotide sequence encoding Hs-unc-53/2 and having the sequence of variant 3 illustrated in Fig. 1c.

Seq ID No. 12 is a nucleotide sequence encoding Hs-unc-53/2 and having the sequence of variant 1 illustrated in Fig. 1c. and lacking the nucleotides from position 5425 to 5433 of the sequence illustrated in Fig. 1c.

Seq ID No. 13 is a nucleotide sequence encoding Hs-unc-53/2 and having the sequence of variant 1 illustrated in Fig. 1c. and lacking the nucleotides from position 5924 to 6024 of the sequence illustrated in Fig. 1c.

Seq ID No. 14 is a nucleotide sequence encoding Hs-unc-53/2 and having the sequence of variant 2 illustrated in Fig. 1c. and lacking the nucleotides from position 5425 to 5433 of the sequence illustrated in Fig. 1c.

Seq ID No. 15 is a nucleotide sequence encoding Hs-unc-53/2 and having the sequence of variant 2 illustrated in Fig. 1c. and lacking the nucleotides

from position 5924 to 6024 of the sequence illustrated in Fig. 1c.

Seq ID No. 16 is a nucleotide sequence encoding Hs-unc-53/2 and having the sequence of variant 3 illustrated in Fig. 1c. and lacking the nucleotides from position 5425 to 5433 of the sequence illustrated in Fig. 1c.

Seq ID No. 17 is a nucleotide sequence encoding Hs-unc-53/2 and having the sequence of variant 3 illustrated in Fig. 1c. and lacking the nucleotides from position 5924 to 6024 of the sequence illustrated in Fig. 1c.

Seq ID No. 18 is an amino acid sequence of Hs-unc-53/2 protein and lacking the amino acids from position 1776 to 1778 of the sequence identified in Fig. 1d

Seq Id No. 19 is an amino acid sequence of variant 1 of Hs-unc-53/2 sequence illustrated in Fig. 1d.

Seq Id No. 20 is an amino acid sequence of variant 2 of Hs-unc-53/2 sequence illustrated in Fig. 1d.

Seq Id No. 21 is an amino acid sequence of variant 3 of Hs-unc-53/2 sequence illustrated in Fig. 1d.

Seq Id No. 22 is an amino acid sequence of variant 1 of Hs-unc-53/2 sequence illustrated in Fig. 1d and lacking the amino acids from position 1776 to 1778 of the sequence identified in Fig. 1d.

Seq Id No. 23 is an amino acid sequence of variant 2 of Hs-unc-53/2 sequence illustrated in Fig. 1d and lacking the amino acids from position 1776 to 1778 of the sequence identified in Fig. 1d.

Seq Id No. 24 is an amino acid sequence of variant 3 of Hs-unc-53/2 sequence illustrated in Fig. 1d and lacking the amino acids from position 1776 to 1778 of the sequence identified in Fig. 1d.

Seq ID No. 25 is a nucleotide sequence encoding Hs-unc-53/3 as illustrated in Figure 1e.

Seq ID No. 26 is a nucleotide sequence encoding Hs-unc-53/3 as illustrated in Figure 1e and lacking the nucleotides from position 3795 to 4283 of the sequence identified therein.

Seq ID No. 27 is a nucleotide sequence encoding Hs-unc-53/3 as illustrated in Figure 1e and lacking the nucleotides from position 4284 to 4325 of the sequence identified therein.

Seq ID No. 28 is a nucleotide sequence encoding Hs-unc-53/3 as illustrated in Figure 1e and lacking the nucleotides from position 3795 to 4325 of the sequence identified therein.

Seq ID No. 29 is a nucleotide sequence encoding Hs-unc-53/3 as illustrated in Figure 1e and lacking the nucleotides from position 5153 to 5173 of the sequence identified.

Seq ID No. 30 is a nucleotide sequence encoding Hs-unc-53/3 as illustrated in Figure 1e and lacking the nucleotides from position 5343 to 5408 of the sequence identified.

Seq ID No. 31 is a nucleotide sequence encoding Hs-unc-53/3 having the sequence of variant 1 illustrated in Fig. 1e.

Seq ID No. 32 is a nucleotide sequence encoding Hs-unc-53/3 having the sequence of variant 1 illustrated in Fig. 1e and lacking the nucleotides from position 3795 to 4283 of the sequence identified therein.

Seq ID No. 33 is a nucleotide sequence encoding Hs-unc-53/3 having the sequence of variant 1 illustrated in Fig. 1e and lacking the nucleotides from position 4284 to 4325 of the sequence identified therein.

Seq ID No. 34 is a nucleotide sequence encoding Hs-unc-53/3 having the sequence of variant 1 illustrated in Fig. 1e and lacking the nucleotides from position 3795 to 4325 of the sequence identified therein.

Seq ID No. 35 is a nucleotide sequence encoding Hs-unc-53/3 having the sequence of variant 1 illustrated in Fig. 1e and lacking the nucleotides from position 5153 to 5173 of the sequence identified therein.

Seq ID No. 36 is a nucleotide sequence encoding Hs-unc-53/3 having the sequence of variant 1 illustrated in Fig. 1e and lacking the nucleotides from position 5343 to 5408 of the sequence identified therein.

Seq ID No. 37 is an amino acid sequence of Hs-unc-53/3 protein as identified in the sequence of Fig. 1f.

Seq ID No. 38 is an amino acid sequence of Hs-unc-53/3 protein as identified in the sequence of Fig. 1f and lacking the amino acid residues from position 1326 to 1413 of the sequence identified therein.

Seq ID No. 39 is an amino acid sequence of Hs-unc-53/3 protein as identified in the sequence of Fig. 1f and lacking the amino acid residues from position 1414

to 1427 of the sequence identified therein.

Seq ID No. 40 is an amino acid sequence of Hs-unc-53/3 protein as identified in the sequence of Fig. 1f and lacking the amino acid residues from position 1703 to 1709 of the sequence identified therein.

Seq ID No. 41 is an amino acid sequence of Hs-unc-53/3 protein as identified in the sequence of Fig. 1f and lacking the amino acid residues from position 1768 to 1788 of the sequence identified therein.

Seq ID No. 42 is an amino acid sequence of Hs-unc-53 of variant 1 identified in Figure 1f.

Seq ID No. 43 is an amino acid sequence of Hs-unc-53 of variant 1 identified in Figure 1f and lacking the amino acid residues from position 1326 to 1413 of the sequence identified therein.

Seq ID No. 44 is an amino acid sequence of Hs-unc-53 of variant 1 identified in Figure 1f and lacking the amino acid residues from position 1414 to 1427 of the sequence identified therein.

Seq ID No. 45 is an amino acid sequence of Hs-unc-53 of variant 1 identified in Figure 1f and lacking the amino acid residues from position 1703 to 1709 of the sequence identified therein.

Seq ID No. 46 is an amino acid sequence of Hs-unc-53 of variant 1 identified in Figure 1f and lacking the amino acid residues from position 1768 to 1788 of the sequence identified therein.

CLAIMS

1. A vertebrate protein homologue of a UNC-53 protein of C. elegans, which protein comprises an amino acid sequence having one or more of sequence blocks A, B, C, D, E, F, G, or H as illustrated in figure 4 or which differs from said blocks in conservative amino acid changes.

2. A vertebrate protein homologue of UNC-53 protein of C. elegans or a functional equivalent, derivative or bioprecursor therefor having an amino acid sequence encoded by the nucleotide sequence illustrated in figure 1(e) or the sequence of Figure 1 e having nucleotide region from position 1 to 288 replaced with the sequence of variant 1 illustrated in Figure 1e and or which sequences further lack any of the sequences form 3795 to 4283, 4284 to 4325, 5153 to 5173 or 5343 to 5408.

3. A vertebrate protein homologue of UNC-53 protein of C. elegans having an amino acid sequence as illustrated in figure 1(f) or an amino acid sequence which differs from said amino acid sequence illustrated in figure 1(f) by the replacement of amino acids 1 to 81 with the sequence of variant 1 in figure 1f and /or including deletions from position 1326 to 1413, 1414 to 1427, 1703 to 1709 or 1768 to 1788, or which differs from said sequences in one or more conservative amino acid changes.

4. A cDNA molecule encoding a vertebrate homologue of UNC-53 protein of C. elegans according to any of claims 1 to 3.

5. A cDNA molecule according to claim 4 which cDNA comprises the sequence of nucleotides illustrated

in figure 1(e).

6. A nucleic acid molecule capable of hybridising to the cDNA sequences according to claims 4 or 5 under high stringency conditions.

7. A DNA expression vector which comprises a cDNA molecule as claimed in claim 4 or 5.

8. A vector according to claim 7 which comprises a promoter of C. elegans UNC-53 protein or a vertebrate homologue thereof according to any of claims 1 to 7.

9. A vector according to claim 8 wherein said promoter sequence is derived from a gene encoding a mouse or human homologue of a UNC-53 protein of C. elegans.

10. A vector according to any of claims 7 to 9 which further comprises a sequence encoding a reporter molecule.

11. A vector according to claim 10 wherein said reporter molecule is a fluorophore.

12. A host cell transformed or transfected with the vector of any of claims 7 to 11.

13. A host cell transformed or transfected with the vector of claims 10 or 11.

14. A host cell according to claim 12 or 13 which cell comprises a prokaryotic cell, such as a bacterial cell or a eukaryotic cell such as a fungal, and animal, a plant or an insect cell.

15. A transgenic cell, tissue or organism comprising a transgene capable of expressing a protein according to any of claims 1 to 3.

16. A transgenic cell, tissue or organism according to claim 15 which comprises any of a COS cell, Hep G2, MCF-7 cell, N4 mouse neuroblastoma cell, a NIH3T£ cell, or colorectal carcinoma or human derived cells.

17. A transgenic cell, tissue or organism according to claim 15 or 16 wherein said transgene comprises a vector according to any of claims 7 to 11.

18. A transgenic cell, tissue or organism according to claim 15 or 17 wherein said transgene comprises a vector according to claim 10 or 11.

19. A transgenic cell, tissue or organism according to any of claims 15 to 17 wherein said organism comprises any of an insect, a fungus, a non-human mammal, a plant or a nematode worm.

20. A method of producing a mutant vertebrate non-human organism which mutation affects cell behaviour or the regulation of cell motility or the shape or the direction of cell migration, which method comprises inducing a mutation in the wild type gene encoding the vertebrate homologue of an UNC-53 C. elegans protein.

21. A vertebrate protein homologue of an UNC-53 protein of C. elegans, according to any of claims 1 to 3 for use as a medicament.

22. Use of a vertebrate protein homologue of an UNC-53 protein of C. elegans, according to any of

claims 1 to 3 in the manufacture of a medicament for promoting neuronal regeneration, revascularisation, wound healing or for treatment of chronic neurodegenerative diseases or acute traumatic injuries or fibrotic disease or autoimmune diseases such as rheumatoid arthritis and sclerosis.

23. A pharmaceutical composition comprising a vertebrate homologue of an UNC-53 protein of C. elegans, according to any of claims 1 to 3 together with a pharmaceutically acceptable carrier, diluent or excipient therefor.

24. A nucleic acid or cDNA molecule according to any of claims 4 to 6 or a functional fragment thereof for use as a medicament.

25. Use of nucleic acid or cDNA molecule according to any of claims 4 to 6 in the manufacture of a medicament to promote neuronal regeneration, revascularisation or wound healing, or for treatment of chronic neurodegenerative diseases or acute traumatic injuries or fibrotic disease or autoimmune diseases such as rheumatoid arthritis and sclerosis.

26. A pharmaceutical composition comprising a nucleic acid or cDNA molecule according to any of claims 4 to 6 and a pharmaceutically acceptable carrier, diluent or excipient therefor.

27. A method of determining whether a compound is an inhibitor or enhancer of the regulation of cell behaviour, growth, cell shape or motility or the direction of cell migration, which method comprises contacting said compound with a host cell according to claim 12 or 14 or a transgenic cell as claimed in any of claims 15 to 18 and screening for a phenotypic

change in said cell.

28. A method according to claim 27 wherein said phenotypic change to be screened is a change in cell growth, or shape or a change in cell motility or filopodia outgrowth, ruffling behaviour, cell adhesion, contact inhibition or the length of neurite growth.

29. A method as claimed in claim 27 wherein said transgenic cell is an N4 neuroblastoma cell and the phenotypic change to the screened is the length of neurite growth.

30. A method as claimed in claim 27 wherein said transgenic cell is an MCF-7 breast carcinoma cell or an NIH3T3 cell and the phenotypic change to be screened is the extent of phagokinesis or contact inhibition.

31. A method of determining whether a compound is an inhibitor or an enhancer of the regulation of cell shape, cell growth or motility or of the direction of cell migration, which method comprises administering said compound to a transgenic organism according to any of claims 15 to 19 or a mutant organism produced according to the method of claim 20 and screening for a phenotypic change in said organism.

32. A compound which is identifiable by the method according to claim 27 as an enhancer of the regulation of cell shape, or growth or motility or the direction of cell migration for use as a medicament.

33. Use of a compound which is identifiable by the method according to claim 27 as an enhancer of the

regulation of cell shape, or growth or motility or the direction of cell migration in the preparation of medicament for promoting neuronal regeneration, revascularisation or wound healing or for treatment of chronic neurodegenerative diseases or acute traumatic injuries or fibrotic disease autoimmune diseases such as rheumatoid arthritis or sclerosis.

34. A pharmaceutical composition comprising a compound identified according to the method of any of claims 27 to 31 and a pharmaceutically acceptable carrier, diluent or excipient therefor.

35. A compound which is identifiable by the method according to any one of claims 17 to 31 as an inhibitor of the regulation of cell motility, growth, or shape, or the direction of cell migration, for use as a medicament.

36. Use of a compound according to claim 35 in the manufacture of a medicament for alleviating the spread of disease inducing cells or metastasis or loss of contact inhibition.

37. A pharmaceutical composition comprising the compound as claimed in claim 35, and a pharmaceutically acceptable carrier diluent or excipient therefor.

38. A method of determining whether a compound is an inhibitor or an enhancer of transcription of a gene encoding a vertebrate homologue of UNC-53 protein of C. elegans, according to any of claims 1 ro 3 which method comprises the steps of (a) contacting said compound with a cell according to claim 13 or 18 and (b) monitoring the level of said reporter molecule and comparing the results obtained from said monitoring

step with a control comprising a cell according to claims 13 or 18, which cell has not been contacted with said compound.

39. A method as claimed in claim 38 wherein said reporter molecule detected is mRNA or green fluorescent protein.

40. A compound which is identifiable by the method according to claims 38 or 39, as an enhancer of transcription of a gene coding for a vertebrate homologue of an UNC-53 protein of C. elegans according to any of claims 1 to 3 or a functional fragment of said gene, for use as a medicament.

41. Use of a compound which is identifiable by the method of claims 38 or 39, as an enhancer of transcription of a gene coding for a vertebrate homologue of an UNC-53 protein of C. elegans according to any of claims 1 to 3 or a functional fragment of said gene, in the manufacture of a medicament for promoting neuronal regeneration, revascularisation or wound healing, or for treatment of chronic neuro-degenerative diseases or acute traumatic injuries or fibriotic disease or autoimmune diseases such as rheumatoid arthritis or sclerosis.

42. A pharmaceutical composition which comprises the compound of claim 40 and a pharmaceutically acceptable carrier, diluent or excipient therefor.

43. A compound which is identifiable by the method of claims 38 or 29 as an inhibitor of transcription of a gene coding for vertebrate homologue of a UNC-53 protein of C. elegans according to any of claims 1 to 3 or a functional fragment of said gene for use as a medicament.

44. Use of a compound which is identifiable by the method of claims 38 or 39 as an inhibitor of transcription of a gene coding for a vertebrate homologue of an UNC-53 protein of C. elegans or a functional fragment of said gene, in the manufacture of a medicament for alleviating spread of disease inducing cells or metastasis or loss of contact inhibition.

45. A pharmaceutical composition which comprises the compound of claim 43 and a pharmaceutically acceptable carrier, diluent or excipient therefor.

46. A kit for determining whether a compound is an enhancer or an inhibitor of the regulation of cell motility, growth or shape or the direction of cell migration which kit comprises at least one transgenic cell as claimed in any one of claims 13 to 17 to be contacted with said compound and at least one cell according to claims 1 2to 19 to be used as a control and means for contacting said compound with one of said at lest one transgenic cells.

47. A kit for determining whether a compound is an inhibitor or an enhancer of transcription of a gene coding for a vertebrate homologue of an UNC-53 protein of C. elegans or a functional fragment of said gene which kit comprises at least one cell as claimed in any one of claims 12 to 19 and means for contacting said compound with said cells.

48. A kit for determining whether a compound is an enhancer or an inhibitor of the activity of a vertebrate homologue of an UNC-53 protein of C. elegans or a functional equivalent, derivative, fragment or bioprecursor of said vertebrate homologue protein, which kit comprises at least, one vertebrate

mutant non-human organism produced according to the method as claimed in claim 20 or a transgenic organism as claimed in claims 15 to 19 and a wild type of said vertebrate mutant organism.

49. A method identifying vertebrate homologues of an unc-53 gene of C. elegans or a functional fragment thereof, which method comprises hybridizing to a DNA library a suitable oligonucleotide sequence of between 15 to 50 nucleotides of the nucleic acid sequence encoding UNC-53 or a functional equivalent, derivative or bioprecursor thereof, under appropriate conditions of stringency to identify genes having statistically significant homology with the cDNA according to any of claims 4 or 5.

50. A method of identifying a protein which is active in the signal transduction pathway of a cell of which a vertebrate homologue of an UNC-53 protein of C. elegans according to any of claims 1 to 3 is a component, which method comprises:

(a) contacting an extract of said cell with an antibody to the vertebrate homologue of the UNC-53 protein of C. elegans,

(b) identifying the antibody/vertebrate homologue complex, and

(c) analysing the complex to identify any protein bound to the vertebrate homologue of UNC-53 protein of C. elegans other than the antibody.

51. A method of identifying a further protein which is active in the signal transduction pathway of a cell of which a vertebrate homologue of an UNC-53 protein according to any of claims 1 to 3 is a component, which method comprises:

(a) forming an antibody to the first identified protein bound to the vertebrate homologue of UNC-53 protein of C. elegans in claim 50,

(b) contacting a cell extract with said antibody and identifying the antibody/protein complex,

(c) analysing the complex to identify any further protein bound to the first protein other than the antibody, and

(d) optionally repeating steps (a) to (c) to identify further proteins in said pathway.

52. A method of identifying a protein which is active in the signal transduction pathway of a cell of which a vertebrate homologue of an UNC-53 protein of C. elegans according to any of claims 1 to 3 is a component, which method comprises:

(a) contacting an extract of said cell with said vertebrate homologue of an UNC-53 protein of C. elegans,

(b) identifying any vertebrate homologue of UNC-53 protein/protein complex formed and

(c) analysing the complex to identify any protein bound to the vertebrate homologue of UNC-53 protein other than the same vertebrate homologue of UNC-53 protein.

53. A method according to claim 52 which further comprises contacting a cell extract with any protein identified from step (c) not being the same as the vertebrate homologue of UNC-53 protein used and repeating steps (b) and (c) so as to identify any further protein involved in the signal transduction pathway of said cell.

54. A method of identifying a protein involved in the signal transduction pathway of a cell of which a vertebrate homologue of an UNC-53 protein of C. elegans is a component which method comprises:

(a) providing an appropriate host cell having a DNA construct comprising a reporter gene under the control of a promoter regulated by a transcription factor having a DNA binding domain and an activating domain,

(b) expressing in said host cell a first hybrid DNA sequence encoding a first fusion of a fragment or all of a DNA sequence according to claims 4 or 5 and either said DNA binding domain or the activating domain of the transcription factor,

(c) expressing in the host cell at least one second hybrid DNA sequence encoding a putative binding protein to be investigated together with the DNA binding or activating domain of the transcription factor which is not incorporated in the first fusion,

(d) detecting any binding of the protein being investigated with a protein according to any of claims 1 to 3 by detecting for the production of any reporter gene product in said host.

55. A protein identified by the method of any one of claims 50 to 54 for use as a medicament.

56. Use of a protein identified by the methods of any one of claims 50 to 54 in the manufacture of a medicament for promoting neuronal regeneration, revascularisation or wound healing, or for treatment of chronic neurodegenerative diseases or acute traumatic injuries or fibrotic disease or autoimmune diseases such as rheumatoid arthritis and sclerosis.

57. A pharmaceutical composition comprising a protein identified by the methods of any one of claims 50 to 54 and a pharmaceutically acceptable carrier, diluent, or excipient therefor.

58. A process for producing a vertebrate homologue of an UNC-53 protein of C. elegans according to any of claims 1 to 3 which process comprises culturing the cells of any of claims 12 to 14 and recovering said vertebrate homologue of UNC-53 protein expressed.

59. A process for producing a vertebrate homologue of an UNC-53 protein of C. elegans according to any of claims 1 to 3 which process comprises culturing an insect cell transfected with a recombinant Baculovirus vector, said vector comprising a DNA insert encoding said vertebrate homologue of UNC-53 protein downstream of the Baculovirus polyhedrin promoter, and recovering the expressed vertebrate homologue of UNC-53 protein.

60. A method of detecting whether a compound is an inhibitor or an enhancer of expression of a vertebrate homologue of an UNC-53 of C. elegans according to any of claims 1 to 3 which method comprises contacting a cell expressing said homologue with said compound and monitoring for a phenotypic change compared to a control cell which has not been contacted with said compound.

61. A method according to claim 60 wherein said cell comprises a cell according to any of claims 12 to 19.

62. A method according to claim 60 wherein said cell has undergone loss of contact inhibition.

63. A method according to any of claims 60 to 62 in which the compound to be tested comprises a nucleic acid.

64. A method according to claim 63 wherein said nucleic acid sequence comprises an antisense DNA or RNA sequence.

65. A method according to claim 64 wherein said mRNA sequence comprises 3′ untranslated regions of mRNA encoding for said vertebrate homologue.

66. A method according to any of claims 60 to 62 wherein said compound to be tested comprises a protein having an amino acid sequence potentially suitable for inhibiting function of said vertebrate homologue.

67. A method according to claim 66 wherein said protein comprises a protein identified according to any of the methods of claims 50 to 54.

68. A pharmaceutical composition comprising a compound identified according to any of claims 60 to 67 together with a pharmaceutically acceptable carrier, diluent or excipient therefor.

69. A nucleic acid sequence identified according to the method of any of claims 63 to 65 for use as a medicament.

70. Use of a nucleotide sequence identified according to the method of any one of claims 63 to 65 in the preparation of a medicament for the treatment of loss of contact inhibition or cancer which is mediated by a vertebrate homologue of an UNC-53 protein of C. elegans.

71. Use of a nucleic acid according to claim 69 in the preparation of a medicament for inhibiting expression of a gene coding for a vertebrate homologue of an UNC-53 protein of C. elegans.

72. An assay for detecting expression of a vertebrate homologue of UNC-53 protein of C. elegans according to any of claims 1 to 3 in a vertebrate cell which assay comprises contacting a cell or an extract thereof with an antibody to said vertebrate homologue, which antibody is linked to a reporter molecule, removing any unbound antibody and monitoring for the presence of said reporter molecule.

73. An assay according to claim 72 wherein said reporter molecule is an antibody conjugated with a suitable fluorophore or detectable enzyme.

74. A method for detecting for expression of a gene coding for a vertebrate homologue of an UNC-53 protein of C. elegans according to any of claims 1 to 3 which method comprises contacting a probe specific for a nucleic acid or protein sequence coding for or corresponding to said vertebrate homologue according to any of claims 1 to 3 with a cell extract which probe is linked to a reporter and analysing for the presence of said reporter.

75. A method according to claim 74 wherein said probe comprises a complementary sequence to a region of mRNA transcribed from said gene encoding said vertebrate homologue of UNC-53 protein.

76. A method according to claim 75 wherein said complimentary sequence is a 3′ or 5′ untranslated region of said mRNA.

77. A method according to claims 74 or 76 wherein said reporter comprises a radiolabel.

78. A method according to claim 74 wherein said probe comprises an antibody specific for said vertebrate homologue of said UNC-53 protein according to any of claims 1 to 3.

79. A method according to claim 78 wherein said reporter comprises an antibody conjugated with a detectable fluorophore or enzyme.

80. A method of determining whether a compound is an inhibitor or an enhancer of association of a vertebrate homologue according to any of claims 1 to 3 to microtubules or plus end regions thereof, which method comprises:-

(a) contacting said compound with a transgenic cell, tissue or organism expressing UNC-53 protein or said vertebrate homologue and which protein is operably linked to a reporter molecule,

(b) screening for the localisation of said reporter molecule as compared to a cell according to step (a) which has not been contacted with said compound.

81. A compound identifiable by the method according to claim 80.

82. A compound according to claim 81 for use as a medicament.

83. Use of a compound according to claim 81 as an enhancer of association of said vertebrate homologue with microtubules or the plus end region thereof, for use in promoting neuronal regeneration,

revascularisation or wound healing, or for treating chronic neurodegenerative diseases or acute traumatic injuries or fibrotic disease or autoimmune diseases such as rheumatoid arthritis or sclerosis.

84. A pharmaceutical composition comprising the compound according to claims 81 or 82 and a pharmaceutically acceptable carrier, diluent or excipient therefor.

85. A kit for determining whether a compound is an inhibitor or an enhancer of association of a vertebrate homologue according to any of claims 1 to 3 with microtubules or the plus end regions thereof, which kit comprises at least one transgenic cell expressing said homologue and a reporter molecule or a cell according to any of claims 12 to 19 and at least one cell of the same cell type for use as a control and means for contacting said compound with one of said at least one transgenic cells.

86. A composition comprising a vertebrate homologue according to any of claims 1 to 3 linked to a compound identified as an inhibitor or enhancer or association of said vertebrate homologue with microtubules or their plus end regions for use in targeting said compound to said microtubule or the plus end region thereof.

87. A composition according to claim 86 which further comprises a cell transformation or transfecting agent.

88. A method of targeting a protein to a cell microtubule or the plus end region thereof, which method comprises introducing into a host cell, tissue or organism a transgene comprising a sequence capable

of expressing a vertebrate homologue according to any of claims 1 to 3, which sequence is operably linked to a sequence encoding said protein to be targeted such that a chimeric protein is expressed and which results in targeting said protein to said microtubule or a plus end region thereof.

89. A method of identifying a molecule which covalently modifies a vertebrate homologue of UNC-53 according to any of claims 1 to 3 which method comprises:

a) contacting an extract from a cell expressing said vertebrate homologue with a mixture of enzymes comprising candidate modifying enzymes in the presence of an inhibitor or covalent modification of a protein,

b) identifying any covalently modified UNC-53 protein from step a),

c) identifying said molecule involved in said modification step.

90. A method according to claim 89, wherein said indicator comprises ^{32}p.

91. A method of identifying a compound which alleviates or enhances the toxicity of a vertebrate homologue according to any of claims 1 to 3, which method comprises contacting said compound with a cell, tissue or organism according to claim 18, and monitoring for the presence of said reporter molecule adjacent said microtubules or the plus end regions thereof.

92. A vertebrate homologue of UNC-53 protein of C.elegans or a functional equivalent, derivative or bioprecursor therefor encoded by the nucleotide sequence in Figure 1a and which nucleotide sequence is

lacking in any of the nucleotide regions from position 2873 to 3043, 3098 to 3121 or 3518 to 3526.

93. A vertebrate homologue of UNC-53 protein of C.elegans or a functional equivalent, derivative or bioprecursor therefor having an amino acid sequence as illustrated in Figure 1b and lacking in one or more of the regions from residues 958 to 1014, 1033 to 1040 or 1173 to 1175, or which differs from said amino acid sequences in one or more conservative amino acid changes.

94. A vertebrate homologue of UNC-53 protein of C.elegans or a functional equivalent, derivative or bioprecursor therefor encoded by the nucleotide sequence in Figure 1c and which nucleotide sequence has from sequence position 1 to 366 replaced with any of the sequences identified as variants 1 to 3 of Figure 1c and/or which sequences lack the region from position 5624 to 6024.

95. A vertebrate homologue of UNC-53 protein of C.elegans or a functional equivalent, derivative or bioprecursor therefor having an amino acid sequence identified in Figure 1d or the sequences of any of variants 1 to 3 replacing the amino acids from position 1 to 89 of the sequence of Figure 1d and/or which sequence is lacking the amino acid sequence from position 1776 to 1778.

96. Plasmid pG313303 deposited under accession number LMBP 3936.

97. Plasmid pG13305 deposited under accession number LMBP 3937.

Figure 1a. Nucleotide sequence of Hs-unc-53/1

```
CATGCTGCCCAAGCGCGCCAAGGCGCCCGGCGGCGGCGGCGGCATGGCCAAGGCCAGCGCGGCTGAGCTGAAGGT   75
CTTCAAGTCCGGCAGCGTGGACAGCCGTGTCCCCGGCGGGCCGCCCGCCTCCAACCTGCGCAAGCAGAAGTCACT  150
CACCAACCTCTCTTTTCTCACGGACTCCGAGAAAAAGCTGCAGCTTTATGAGCCCGAATGGAGCGACGATATGGC  225
CAAGGCGCCCAAAGGCTTAGGCAAGGTGGGGTCCAAGGGCCGTGAAGCTCCGCTGATGTCCAAGACGCTGTCCAA  300
GTCGGAGCACTCGCTCTTCCAGGCCAAGGGCAGCCCGGCGGGCGGTGCCAAGACCCCCCTGGCTCCGCTCGCGCC  375
CAACCTGGGAAAGCCGAGCCGGATCCCTCGAGGACCCTATGCGGAGGTCAAGCCGCTCAGCAAGGCGCCTGAAGC  450
GGCCGTGAGCGAAGATGGCAAATCGGACGACGAGCTGCTCTCCAGCAAGGCCAAGGCGCAAAAGAGCTCTGGGCC  525
TGTCCCCTCTGCCAAGGGCCAGGAGGAGCGCGCCTTCCTCAAGGTGGACCCCGAGCTGGTGGTGACCGTGCTGGG  600
AGACCTGGAGCAGCTGCTCTTCAGCCAGATGCTGGACCCAGAGTCCCAGAGAAAGAGGACAGTGCAGAATGTCCT  675
GGATCTCCGGCAGAACCTGGAAGAGACCATGTCCAGCCTGCGAGGGTCCCAGGTGACTCACAGCTCCCTGGAGAT  750
GACCTGCTACGACAGCGATGATGCCAACCCACGCAGCGTGTCCAGCCTCTCCAACCGCTCGTCCCCTCTGTCATG  825
GCGCTATGGCCAGTCCAGTCCGCGGCTGCAGGCTGGTGACGCGCCCTCTGTGGGTGGGAGCTGCCGCTCGGAGGG  900
GACGCCCGCCTGGTACATGCACGGCGAACGGGCCCACTACTCCCACACCATGCCCATGCGCAGCCCCAGCAAGCT  975
CAGCCATATCTCCCGCCTGGAGCTGGTCGAATCCCTGGACTCGGATGAGGTGGACCTCAAGTCCGGCTACATGAG 1050
CGACAGTGACCTCATGGGCAAGACCATGACGGAGGATGATGACATCACTACCGGCTGGGATGAAAGCAGCTCCAT 1125
CAGTAGTGGACTCAGCGATGCCTCAGACAATCTCAGTTCAGAAGAATTCAATGCCAGCTCCTCACTCAACTCCCT 1200
CCCAAGTACTCCCACTGCTTCTCGCAGGAACTCAACAATAGTGCTACGCACAGACTCAGAGAAGCGCTCACTGGC 1275
AGAAAGTGGGCTGAGCTGGTTTAGTGAATCAGAGGAGAAAGCCCCTAAAAAACTGGAGTACGACAGTGGTAGCCT 1350
GAAGATGGAACCTGGGACTTCTAAGTGGCGGAGGGAGCGGCCTGAGAGCTGTGATGATTCATCCAAGGGTGGAGA 1425
ACTGAAAAAGCCCATCAGCCTGGGCCACCCTGGTTCCCTGAAGAAGGGCAAGACCCCACCTGTGGCTGTAACTTC 1500
CCCCATCACTCACACAGCCCAGAGTGCCCTCAAAGTCGCAGGCAAACCTGAGGGCAAAGCTACAGACAAGGGTAA 1575
GCTTGCAGTGAAGAATACTGGGCTCCAACGCTCCTCCTCTGATGCTGGTCGGGACCGCCTGAGTGATGCTAAGAA 1650
GCCCCCCTCGGGCATTGCTCGCCCCTCCACTTCGGGATCCTTTGGCTACAAGAAGCCTCCTCCTGCCACAGGCAC 1725
AGCCACTGTCATGCAAACTGGTGGTTCAGCCACTCTCAGCAAGATCCAGAAGTCCTCAGGCATCCCTGTCAAGCC 1800
AGTAAATGGGCGCAAGACTAGCTTAGATGTTTCCAACAGTGCAGAGCCAGGATTCCTGGCTCCTGGAGCCCGTTC 1875
TAACATCCAGTACCGCAGCCTGCCCCGGCCAGCCAAGTCAAGTTCTATGAGCGTGACCGGCGGGCGGGGTGGACC 1950
TCGCCCTGTGAGCAGCAGCATTGACCCCAGTCTCCTCAGCACCAAGCAGGGAGGCCTTACGCCTTCCAGACTGAA 2025
GGAGCCTACCAAGGTAGCCAGTGGGCGGACCACTCCAGCCCCTGTCAATCAGACAGATCGGGAAAAGGAGAAGGC 2100
CAAAGCCAAGGCAGTGGCCTTGGACTCAGACAACATCTCCTTGAAGAGTATTGGCTCCCCAGAAAGTACTCCCAA 2175
GAACCAAGCAAGCCACCCCACAGCCACCAAGCTGGCAGAGCTGCCACCAACCCCTCTCAGGGCCACAGCGAAGAG 2250
CTTTGTCAAACCACCCTCACTAGCCAATCTTGACAAGGTCAACTCCAACAGTCTGGATCTACCATCATCCAGTGA 2325
TACCACCCATGCTTCAAAGGTCCCAGATCTGCATGCTACAAGCTCAGCATCTGGGGGCCCTCTCCCTTCCTGCTT 2400
CACCCCCAGTCCGGCACCCATCCTCAATATTAACTCAGCCAGCTTCTCCCAGGGCCTGGAGCTAATGAGTGGTTT 2475
CAGTGTGCCAAAAGAGACCCGCATGTACCCCAAACTCTCAGGCCTGCACAGGAGCATGGAGTCCCTCCAGATGCC 2550
AATGAGCCTCCCCAGTGCCTTCCCCAGCAGTACTCCCGTCCCCACCCCACCTGCTCCCCCTGCTGCTCCCACAGA 2625
AGAAGAGACGGAAGAGCTGACTTGGAGTGGAAGCCCCAGAGCTGGGCAACTGGACAGTAATCAGCGGGATCGGAA 2700
CACTCTTCCCAAGAAAGGGCTCAGGTACCAGCTTCAGTCCCAGGAGGAGACCAAGGAGAGGCGACATTCCCATAC 2775
CATTGGTGGGCTGCCTGAATCCGATGACCAGTCAGAGCTGCCTTCTCCCCCTGCACTTCCCATGTCTCTGAGTGC 2850
AAAGGGCCAACTTACCAACATAgtgagtcccactgcggccaccacgccaagaatcacccgctccaacagcatccc 2925
cacccacgaggcggccttcgagctgtacagcggctcccaaatggggagcaccctgtccctggccgagagacccaa 3000
gggaatgattcggtcaggatccttccgagaccccacggacgatGTTCACGGCTCAGTGCTGTCCCTGGCCTCCAG 3075
TGCCTCCTCCACCTACTCCTCAgctgaggagaggatgcaatctgagCAAATCCGGAAGCTTCGTAGGGAACTGGA 3150
ATCATCCCAGGAAAAAGTGGCCACCTTGACGTCTCAGCTTTCTGCCAATGCTAATCTGGTGGCTGCTTTTGAGCA 3225
GAGCCTGGTGAATATGACATCCCGCCTGCGACACCTGGCAGAGACGGCCGAGGAGAAGGACACTGAGCTGCTGGA 3300
TTTGCGAGAAACCATAGACTTTCTGAAGAAAAAGAACTCTGAGGCCCAGGCAGTCATTCAGGGAGCCCTTAATGC 3375
CTCAGAAACCACACCCAAAGAACTTCGGATCAAGAGACAAAACTCCTCAGATAGCATCTCAAGCCTCAACAGCAT 3450
CACTAGCCATTCCAGCATCGGCAGCAGCAAGGATGCTGATGCGAAAAAGAAGAAAAAAAAGAGTTGGGtctatga 3525
gCTTCGAAGTTCCTTCAACAAAGCGTTCAGTATAAAAAAGGGGCCCAAGTCAGCTTCCTCATACTCGGATATAGA 3600
GGAGATTGCTACACCCGACTCTTCAGCCCCCTCATCCCCCAAACTACAGCATGGTTCTACAGAGACTGCTTCACC 3675
CTCCATCAAGTCCTCCACCTYGTCCTCCGTGGGCACTGATGTCACCGAGGGCCCTGCTCACCCAGCCCCCCACAC 3750
TAGGCTGTTCCATGCAAATGAGGAGGAGGAGCCAGAGAAGAAGGAGGTATCGGAGCTGCGCTCTGAGCTATGGGA 3825
GAAGGAAATGAAGCTTACAGACATCCGCTTGGAGGCCCTCAACTCTGCCCACCAACTGGATCAGCTTCGGGAGAC 3900
CATGCACAACATGCAGTTGGAGGTGGACCTGCTGAAAGCAGAGAATGACCGACTGAAGGTAGCCCCAGGCCCCTC 3975
ATCAGGCTCCACTCCAGGGCAGGTCCCTGGATCATCTGCATTATCTTCCCCACGCCGCTCCCTAGGCCTGGCACT 4050
CACCCATTCCTTCGGCCCCAGTCTTGCAGACACAGACCTGTCACCCATGGATGGCATCAGTACTTGTGGTCCAAA 4125
GGAGGAAGTGACCCTCCGGGTGGTGGTGAGGATGCCCCCGCAGCACATCATCAAAGGGGACTTGAAGCAGCAGGA 4200
ATTCTTCCTGGGCTGTAGCAAGGTCAGTGGAAAAGTTGACTGGAAGATGCTGGATGAAGCTGTTTTCCAAGTGTT 4275
CAAGGACTATATTTCTAAAATGGACCCAGCCTCTACCCTGGGACTAAGCACTGAGTCCATCCATGGCTACAGCAT 4350
CAGCCACGTGAAACGAGTGTTGGATGCAGAGCCCCCCGAGATGCCTCCTTGCCGTCGAGGTGTCAATAACATATC 4425
AGTCTCCCTCAAAGGTCTGAAGGAGAAATGCGTCGACAGCCTGGTGTTCGAGACGCTGATCCCCAAGCCGATGAT 4500
GCAGCACTACATAAGCCTCCTGCTGAAGCACCGGCGCCTCGTCCTCTCGGGCCCCAGCGGCACGGGCAAGACCTA 4575
CCTGACCAATCGCTTGGCCGAGTACCTGGTGGAGCGCTCTGGCCGTGAGGTCACAGAGGGCATCGTCAGCACCTT 4650
CAACATGCACCAGCAGTCTTGCAAGGATCTGCAACTGTATCTTTCCAACCTAGCCAACCAGATAGACCGGGAAAC 4725
AGGAATTGGGGATGTGCCCCTGGTGATTCTATTGGATGACCTGAGTGAAGCAGGCTCCATCAGTGAGTTGGTCAA 4800
TGGGGCCCTCACCTGCAAGTATCATAAATGTCCCTATATTATAGGTACCACCAATCAGCCTGTAAAAATGACACC 4875
```

Figure 1a (CONTINUED)

```
CAACCATGGCTTGCACTTGAGCTTCAGGATGTTGACCTTCTCCAACAACGTGGAGCCAGCCAATGGCTTCCTGGT 4950
TCGTTACCTGAGGAGGAAGCTGGTAGAGTCAGACAGCGACATCAATGCCAACAAGGAAGAGCTGCTTCGGGTGCT 5025
CGACTGGGTACCCAAGCTGTGGTATCATCTCCACACCTTCCTTGAGAAGCACAGCACCTCAGACTTCCTCATCGG 5100
CCCTTGCTTCTTTCTGTCGTGTCCCATTGGCATTGAGGACTTCCGGACCTGGTTCATTGACCTGTGGAACAACTC 5175
TATCATTCCCTATCTACAGGAAGGAGCCAAGGATGGGATAAAGGTCCATGGACAGAAAGCTGCTTGGGAGGACCC 5250
AGTGGAATGGGTCCGGGACACACTTCCCTGGCCATCAGCCCAACAAGACCAATCAAAGCTGTACCACCTGCCCCC 5325
ACCCACCGTGGGCCCTCACAGCATTGCCTCACCTCCCGAGGATAGGACAGTCAAAGACAGCACCCCAAGTTCTCT 5400
GGACTCAGATCCTCTGATGGCCATGCTGCTGAAACTTCAAGAAGCTGCCAACTACATTGAGTCTCCAGATCGAGA 5475
AACCATCCTGGACCCCAACCTTCAGGCAACACTTTAAGGGTTCGGCAATCACTGTCACCCCCGGACAGCAGAACG 5550
CTGGCATCAGCTATCTTAGCTCCTCCTCTCCCCTCTCCTCTTTCAGAGCACTGGCTCTCCAGCCCCAGGAGGAGA 5625
ACAGGAGGGAGGAGGAGATGAAAGAGGAGGGACAGGTTCTTGGTGCTGTACCTTTGAGAACTTCCTAGGAAGGAA 5700
TGGTGGGGTGGCGTTTGGGAACTTGTGCCCCCTAAACACATTTACTGGCCTCCTCTAATGACTTTGGGGAAAAGA 5775
TGATTCTGGGTCTTTCCCTTGACTTCTTGTTTCAATTACAAACTCCTGGGCTTTCTGGGGAGGGGTTCAGAAAAC 5850
ATCAAAACACTGCAGCAGTTCCTAAATGATTCTCACAAGCAACCCTGAGAGAGACAGTCTTGTGAGGGAGATCTG 5925
GGGGAGGCAGGAAGCTCCTCAGATTTTCTCACAGACCCTTCCCAATTCCATCACCACTGCCAACACTCGTCCGGA 6000
ATTC                                                                          6004
```

In frontal cortex, variants have been found lacking the region from position 2873 to 3043 or the region from residues 3098 to 3121. The region from 3518 to 3526 is absent in cDNA from Hela or colorectal adenocarcinoma tissue. All three regions are indicated in lower case letters in the figure above. Y at position 3696 stands for C or T. Both nucleotides have been found to be present in cDNAs from different origin.

Figure 1b. Amino Acid sequence of the protein encoded by Hs-unc-53/1 gene. Stretches encoded by the DNA sequences lacking in variants from frontal cortex are in lower case letters (residues 958 to 1014 ; 1033 to 1040 and 1173 to 1175). The x at position 1232 stands for Leucine or Serine, depending on the cDNA of origin.

```
MLPKRAKAPGGGGGMAKASAAELKVFKSGSVDSRVPGGPPASNLRKQKSLTNLSFLTDSEKKLQLYEPEWSDDMA   75
KAPKGLGKVGSKGREAPLMSKTLSKSEHSLFQAKGSPAGGAKTPLAPLAPNLGKPSRIPRGPYAEVKPLSKAPEA  150
AVSEDGKSDDELLSSKAKAQKSSGPVPSAKGQEERAFLKVDPELVVTVLGDLEQLLFSQMLDPESQRKRTVQNVL  225
DLRQNLEETMSSLRGSQVTHSSLEMTCYDSDDANPRSVSSLSNRSSPLSWRYGQSSPRLQAGDAPSVGGSCRSEG  300
TPAWYMHGERAHYSHTMPMRSPSKLSHISRLELVESLDSDEVDLKSGYMSDSDLMGKTMTEDDDITTGWDESSSI  375
SSGLSDASDNLSSEEFNASSSLNSLPSTPTASRRNSTIVLRTDSEKRSLAESGLSWFSESEEKAPKKLEYDSGSL  450
KMEPGTSKWRRERPESCDDSSKGGELKKPISLGHPGSLKKGKTPPVAVTSPITHTAQSALKVAGKPEGKATDKGK  525
LAVKNTGLQRSSSDAGRDRLSDAKKPPSGIARPSTSGSFGYKKPPPATGTATVMQTGGSATLSKIQKSSGIPVKP  600
VNGRKTSLDVSNSAEPGFLAPGARSNIQYRSLPRPAKSSSMSVTGGRGGPRPVSSSIDPSLLSTKQGGLTPSRLK  675
EPTKVASGRTTPAPVNQTDREKEKAKAKAVALDSDNISLKSIGSPESTPKNQASHPTATKLAELPPTPLRATAKS  750
FVKPPSLANLDKVNSNSLDLPSSSDTTHASKVPDLHATSSASGGPLPSCFTPSPAPILNINSASFSQGLELMSGF  825
SVPKETRMYPKLSGLHRSMESLQMPMSLPSAFPSSTPVPTPPAPPAAPTEEETEELTWSGSPRAGQLDSNQRDRN  900
TLPKKGLRYQLQSQEETKERRHSHTIGGLPESDDQSELPSPPALPMSLSAKGQLTNIvsptaattpritrsnsip  975
theaafelysgsqmgstlslaerpkgmirsgsfrdptddVHGSVLSLASSASSTYSSaeermqseQIRKLRRELE 1050
SSQEKVATLTSQLSANANLVAAFEQSLVNMTSRLRHLAETAEEKDTELLDLRETIDFLKKKNSEAQAVIQGALNA 1125
SETTPKELRIKRQNSSDSISSLNSITSHSSIGSSKDADAKKKKKKSWvyeLRSSFNKAFSIKKGPKSASSYSDIE 1200
EIATPDSSAPSSPKLQHGSTETASPSIKSSTxSSVGTDVTEGPAHPAPHTRLFHANEEEEPEKKEVSELRSELWE 1275
KEMKLTDIRLEALNSAHQLDQLRETMHNMQLEVDLLKAENDRLKVAPGPSSGSTPGQVPGSSALSSPRRSLGLAL 1350
THSFGPSLADTDLSPMDGISTCGPKEEVTLRVVVRMPPQHIIKGDLKQQEFFLGCSKVSGKVDWKMLDEAVFQVF 1425
KDYISKMDPASTLGLSTESIHGYSISHVKRVLDAEPPEMPPCRRGVNNISVSLKGLKEKCVDSLVFETLIPKPMM 1500
QHYISLLLKHRRLVLSGPSGTGKTYLTNRLAEYLVERSGREVTEGIVSTFNMHQQSCKDLQLYLSNLANQIDRET 1575
GIGDVPLVILLDDLSEAGSISELVNGALTCKYHKCPYIIGTTNQPVKMTPNHGLHLSFRMLTFSNNVEPANGFLV 1650
RYLRRKLVESDSDINANKEELLRVLDWVPKLWYHLHTFLEKHSTSDFLIGPCFFLSCPIGIEDFRTWFIDLWNNS 1725
IIPYLQEGAKDGIKVHGQKAAWEDPVEWVRDTLPWPSAQQDQSKLYHLPPPTVGPHSIASPPEDRTVKDSTPSSL 1800
DSDPLMAMLLKLQEAANYIESPDRETILDPNLQATL                                          1835
```

Figure 1c. Nucleotide sequence of the Hs-unc-53/2 2 gene

```
TAAGGCCCGGCGCCTGCTCTGCTACCCGCGCTGCCTTTAGCGGTCGCCCCCGCCGCCGCTGCCAGGGACGTGCTG    75
GGAAAGCCCAAGCCCCGGGAGAAGATGCCGGCCATCCTGGTCGCCTCCAAAATGAAGTCGGGACTGCCCAAACCC   150
GTGCACAGCGCCGCGCCCATCCTGCACGTGCCCCCGGCCCGGGCGGGCCCCCAGCCCTGCTACCTGAAGTTGGGA   225
AGCAAGGTGGAGGTGAGCAAGACCACCTATCCTAGCCAGATCCCCCTGAAATCGCAGGTGCTGCAGGGGCTGCAG   300
GAGCCAGCGGGGGAGGGGCTCCCGCTGCGGAAGAGCGGCTCGGTGGAAAACGGGTTCGATACCCAGATCTACACA   375
GACTGGGCCAATCATTACCTAACCAAATCCGGCCACAAGCGTCTCATCAAGGATCTCCAGCAAGATGTGACAGAT   450
GGCGTCCTCCTGGCCCAGATTATCCAGGTTGTGGCAAATGAAAAGATTGAAGACATCAATGGCTGTCCGAAGAAC   525
AGATCCCAAATGATTGAAAACATAGATGCCTGCTTGAATTTCCTGGCAGCTAAGGGAATAAACATCCAGGGGCTG   600
TCTGCAGAAGAGATCAGGAATGGAAACCTCAAGGCCATTCTAGGCCTCTTCTTCAGCCTCTCCCGATACAAGCAG   675
CAGCAGCAGCAGCCCCAGAAGCAGCACCTCTCCTCACCTCTGCCGCCCGCCGTATCCCAGGTGGCCGGGGCCCCC   750
TCCCAGTGCCAGGCTGGCACCCCTCAGCAGCAGGTGCCAGTCACTCCCCAAGCCCCGTGCCAGCCTCACCAGCCA   825
GCGCCACATCAGCAGTCAAAAGCACAAGCTGAAATGCAGTCCAGACTTTCAGGTCCTACCGCGAGGGTATCCGCT   900
GCAGGCAGCGAGGCCAAAACACGCGGAGGGTCAACTACTGCTAACAACCGACGCAGCCAGAGCTTTAACAACTAT   975
GATAAATCCAAACCAGTCACCTCCCCACCCCCACCGCCAAGCAGCCACGAGAAAGAGCCTTTGGCAAGTTCAGCC  1050
TCCTCCCACCCCGGAATGAGTGACAATGCACCTGCTTCCTTGGAGAGCGGCAGCAGCTCCACCCCTACTAATTGC  1125
AGTACCTCCTCGGCCATCCCGCAGCCCGGTGCAGCCACCAAGCCTTGGCGCAGCAAATCCCTCAGCGTGAAGCAC  1200
AGTGCCACGGTATCCATGCTCTCGGTCAAGCCTCCTGGGCCTGAGGCCCCCAGGCCCACACCTGAAGCCATGAAG  1275
CCGGCCCCCAACAATCAGAAGTCCATGCTGGAAAAGCTGAAACTTTTCAACAGTAAAGGGGGCTCAAAGGCAGGT  1350
GAGGGGCCGGGGTCCCGGGACACAAGCTGTGAGCGGCTGGAGACTCTGCCCAGCTTCGAAGAGAGCGAGGAGCTG  1425
GAGGCCGCCAGTCGCATGCTCACCACCGTGGGCCCTGCTTCCAGCAGCCCCAAGATTCCACTCAAGGCCATTGCC  1500
CAGAGGACTTTTAGCCGGGCACTGACCAACAAGAAGAGTTCTCTGAAAGGCAATGAGAAAGAGAAGGAGAAACAA  1575
CAGCGGGAGAAGGATAAGGAGAAAAGCAAGGACCTTGCCAAGAGAGCCTCTGTGACGGAGAGGCTGGACCTCAAG  1650
GAGGAGCCAAAAGAAGACCCCAGTGGAGCAGCTGTGCCCGAGATGCCAAAAAAGTCCTCCAAGATTGCCAGCTTC  1725
ATCCCCAAAGGGGGGAAGCTCAACAGTGCCAAGAAGGAGCCCATGGCCCCTTCCCACAGTGGAATACCAAAACCA  1800
GGAATGAAGAGCATGCCCGGGAAATCCCCAAGTGCCCCAGCGCCTTCCAAGGAAGGGGAGCGGAGCCGGAGTGGG  1875
AAGCTGAGCTCAGGACTCCCCCAGCAGAAGCCCCAGCTGGACGGCAGACACTCCAGTTCCTCTTCCAGCCTGGCG  1950
TCCTCAGAAGGAAAAGGCCCAGGAGGGACCACCCTGAACCACAGCATCAGCAGCCAGACTGTCAGTGGGTCTGTC  2025
GGGACCACCCAGACCACAGGAAGCAATACCGTCAGTGTTCAGCTACCTCAGCCCCAGCAGCAATACAACCATCCC  2100
AACACTGCCACGGTTGCACCTTTCCTGTACAGGTCTCAGACGGACACTGAAGGGAATGTTACTGCCGAGTCAAGC  2175
TCAACAGGTGTGAGCGTGGAGCCCAGCCACTTCACCAAGACTGGACAGCCTGCTCTGGAAGAACTCACTGGGGAA  2250
GATCCTGAGGCTCGGCGGCTGCGGACAGTGAAGAACATCGCTGATCTGCGGCAGAATTTGGAGGAAACCATGTCC  2325
AGTTTAAGGGGAACTCAGGTTACACACAGCACATTGGAAACCACGTTTGACACCAATGTCACCACGGAGATGAGT  2400
GGCCGTAGCATACTCAGCTTGACAGGGAGGCCCACACCTCTGTCCTGGAGACTGGGCCAGTCCAGCCCTCGGCTC  2475
CAAGCAGGAGACGCCCCCTCAATGGGCAATGGGTATCCCCCTCGAGCCAACGCCAGCAGGTTCATCAACACTGAG  2550
TCAGGTCGCTATGTGTACTCCGCCCCTCTGAGAAGGCAGCTGGCCTCCCGGGGCAGTAGTGTCTGCCAYGTGGAC  2625
GTCTCAGACAAGGCAGGAGATGAGATGGACCTGGAAGGCATCAGCATGGACGCCCCCGGCTACATGAGCGATGGG  2700
GATGTTCTGAGCAAGAACATCCGGACCGATGACATTACAAGCGGATACATGACTGATGGTGGACTTGGCCTCTAT  2775
ACCCGTCGCCTGAACCGGCTCCCTGATGGGATGGCTGTGGTACGGGAGACCCTGCAACGAAATACCTCCCTGGGC  2850
CTCGGAGACGCTGACAGCTGGGACGACAGCAGCTCCGTCAGCAGCGGCATCAGCGACACCATAGACAACCTCAGC  2925
ACTGATGACATCAACACCAGCTCCTCCATCAGCTCTTATGCCAACACACCTGCCTCCTCTCGAAAAAACCTGGAT  3000
GTGCAGACTGATGCTGAGAAGCACTCACAGGTGGAGAGGAATTCCCTGTGGTCTGGTGATGATGTCAAGAAATCA  3075
GACGGAGGCTCAGACAGCGGCATAAAAATGGAGCCAGGTTCCAAGTGGAGGCGGAATCCTTCTGATGTGTCTGAC  3150
GAKTCCGACAAAAGCACGTCGGGCAAGAAGAATCCTGTCATCTCCCAGACAGGCTCATGGCGGCGAGGCATGACA  3225
GCTCAGGTGGGCATCACCATGCCAAGGACGAAGGCTTCAGCCCCGGCAGGCGCACTGAAGACCCCAGGAACTGGA  3300
AAAACAGACGACGCAAAGGTGTCTGAGAAAGGAAGGCTTTCTCCTAAAGCCTCCCAGGTGAAGCGCTCCCCATCA  3375
GATGCAGGCCGGAGCAGTGGTGACGAATCCAAAAAGCCCCTCCCCAGCAGCTCTAGGACACCTACTGCCAATGCC  3450
AACAGCTTTGGGTTCAAGAAGCAGAGTGGTTCCGCCGCCGGCCTGGCCATGATCACAGCCAGCGGGGTGACTGTC  3525
ACCAGCAGGTCAGCCACACTGGGCAAAATCCCAAAGTCATCTGCACTCGTCAGTCGGTCTGCTGGTCGGAAGTCA  3600
AGTATGGATGGGGCTCAGAATCAGGATGACGGGTATCTAGCCCTAAGCTCCCGGACAAACCTTCAGTACCGGAGT  3675
TTGCCGAGGCCCAGTAAGTCCAACAGCCGGAACGGGGCTGGGAACAGGTCTAGCACCAGCAGCATAGATTCCAAC  3750
ATTAGCAGCAAGTCCGCAGGCCTGCCAGTGCCCAAACTGAGGGAGCCTTCCAAAACAGCCCTAGGCAGCTCTCTA  3825
CCAGGTCTGGTCAACCAAACAGATAAGGAGAAAGGCATCTCATCAGACAACGAGAGTGTGGCTTCCTGTAACTCG  3900
GTGAAAGTGAATCCGGCAGCCCAGCCTGTGTCCAGTCCGGCTCAGACCAGTCTCCAGCCTGGAGCCAAGTACCCA  3975
GATGTGGCCTCTCCCACACTCCGCAGACTCTTTGGTGGGAAGCCTACCAAGCAAGTGCCCATCGCCACAGCTGAA  4050
AACATGAAAAATTCGGTGGTCATCTCCAATCCTCATGCCACCATGACTCAGCAAGGTAACCTAGACTCCCCGTCA  4125
GGCAGTGGCGTCCTGAGCAGTGGGAGCAGCAGTCCTCTCTACAGCAAGAATGTGGACCTCAACCAGTCTCCGCTA  4200
GCCTCCAGCCCCAGCTCAGCCCACTCGGCCCCTTCCAACAGCCTCACCTGGGGCACCAACGCCAGCAGCTCCTCC  4275
GCAGTTAGCAAGGATGGCCTGGGCTTTCAGTCTGTCAGCAGCCTCCACACCAGCTGTGAGTCCATCGACATCTCC  4350
CTCAGCAGTGGAGGGGTCCCCAGCCACAATTCTTCCACTGGCCTCATCGCCTCCTCCAAGGACGACTCCTTGACT  4425
CCCTTTGTCAGAACTAACAGTGTGAAGACCACACTGTCAGAAAGCCCTCTCTCTTCCCCTGCTGCTAGCCCTAAG  4500
TTCTGCAGAAGTACTCTGCCCAGGAAACAGGACAGTGACCCGCACCTTGATAGGAACACTTTGCCTAAGAAAGGA  4575
CTCAGGTATACTCCCACCTCCCAGCTTCGCACGCAAGAAGATGCAAAAGAATGGTTACGGTCCCATTCTGCAGGA  4650
GGCCTTCAGGACACCGCTGCCAATTCCCCCTTTTCCTCTGGCTCCAGCGTGACTTCTCCCTCCGGAACAAGATTC  4725
AACTTTTCCCAGCTTGCGAGTCCCACCACTGTCACCCAGATGAGCTTGTCCAACCCGACCATGCTGAGGACTCAC  4800
AGCCTCTCCAATGCTGATGGGCAGTATGATCCATACACTGACAGCCGCTTCCGGAATAGCTCCATGTCCCTGGAT  4875
```

WO 99/63080 PCT/EP99/03848

Figure 1C (CONTINUED)

GAGAAGAGCAGAACCATGAGCCGTTCAGGCTCATTCCGGGATGGGTTTGAAGAAGTTCATGGATCCTCACTCTCC 4950
TTGGTTTCCAGCACATCGTCAGTTTATTCTACACCAGAAGAAAAATGCCAGTCAGAGATTCGCAAGCTGCGGCGG 5025
GAACTGGATGCCTCCCAGGAGAAAGTTTCAGCTTTGACCACCCAGCTGACAGCAAAYGCTCACCTTGTGGCWGCC 5100
TTTGAACAGAGTCTTGGTAACATGACAATCAGGCTCCAGAGTCTGACCATGACAGCTGAGCAGAAGGATTCAGAA 5175
CTGAATGAGTTAAGAAAAACCATTGAGCTGCTAAAGAAACAGAACGCAGCTGCCCAGGCTGCCATCCGCAGGCAGCAC 5250
ATTAACACACCTGAGCTCAACTGCAAAGGAAACGGCACTGCCCAGTCTGCAGACCTCCGCATCCGCAGGCAGCAC 5325
TCCTCAGACAGCGTCTCCAGCATCAACAGTGCCACCAGCCACTCCAGTGTGGGCAGCAACATAGAGAGTGACTCA 5400
AAGAAGAAGAAGAGGAAGAACTGGgtcaatgagTTACGCAGCTCCTTCAAGCAAGCTTTCGGGAAGAAGAAGTCC 5475
CCAAAATCTGCGTCCTCTCATTCAGATATTGAGGAGATGACGGATTCTTCTTTGCCTTCCTCACCAAAGTTACCR 5550
CACAATGGGTCCACAGGTTCCACCCCACTGCTGAGGAATTCTCACTCCAACTCTCTAATTTCMGAATGCATGGAT 5625
AGTGAAGCTGAGACCGTCATGCAGCTCCGAAATGAGTTAAGAGACAAGGAGATGAAGCTGACRGATATCCGCTTA 5700
GAAGCTCTCAGTTCTGCCCACCAGCTGGACCAGCTCCGGGAGGCCATGAACAGGATGCAGAGTGAAATAGAGAAG 5775
CTGAAAGCTGAGAATGATCGGCTGAAGTCAGAGTCTCAAGGCAGTGGCTGCAGCCGGGCTCCTTCCCAAGTGTCC 5850
ATCTCTGCCTCCCCGAGGCAGTCCATGGGCCTCTCCCAGCACAGCTTGAACCTCACTGAGTCAACCAGCCTGGAC 5925
ATGTTGCTGGATGACACTGGTGAATGCTCGGCTCGGAAGGAAGGAGGCAGGCATGTTAAGATAGTTGTCAGCTTT 6000
CAGGAGGAAATGAAGTGGAAGGAGGATTCCAGACCACAYCTCTTTCTTATTGGCTGCATTGGAGTTAGTGGCAAG 6075
ACGAAGTGGGATGTGCTCGATGGGGTGGTTAGACGGCTGTTCAAAGAATACATCATTCATGTCGACCCAGTGAGT 6150
CAGCTAGGGCTGAATTCAGACAGCGTTCTTGGCTACAGCATTGGAGAAATCAAGCGCAGCAACACTTCCGAAACA 6225
CCGGAGCTGCTTCCTTGTGGCTATCTGGTTGGAGAGACACGACCATCTCAGTGAAAGGGCTCGCAGAA 6300
AACAGCCTGGACTCACTGGTGTTTGAGTCCTTGATTCCCAAGCCCATCCTGCAGCGCTACGTCTCCCTCCTGATA 6375
GAGCACCGTCGGATCATTCTCTCTGGCCCCAGCGGCACTGGGAAAACCTACCTGGCCAACCGGCTGTCTGAGTAT 6450
ATAGTGCTTCGAGAGGGACGGGAGTTGACAGACGGGGTTATCGCCACCTTTAACGTGGACCATAAGTCCAGCAAG 6525
GAATTGCGCCAGTACCTGTCCAACCTTGCTGACCAGTGCAACAGTGAGAACAATGCTGTGGACATGCCCCTCGTC 6600
ATCATCCTGGACAACCTACACCACGTGAGCTCTCTGGGCGAGATCTTCAATGGGCTGCTCAACTGCAAGTACCAC 6675
AAATGCCCTTACATAATTGGCACAATGAACCAGGCTACCTCTTCGACTCCCAACCTGCAGCTTCACCATAACTTC 6750
AGATGGGTGCTTTGTGCCAACCACACGGAGCCTGTGAAGGGTTTCCTTGGCCGATTCCTGAGGAGGAAGCTCATG 6825
GAAACAGAGATCAGTGGGCGGGTGCGCAATATGGAGCTGGTAAAAATCATTGACTGGATTCCCAAGGTCTGGCAT 6900
CACCTCAACCGCTTCCTGGAGGCTCACAGTTCCTCGGACGTCACCATCGGCCCCCGGCTCTTCCTGTCATGCCCC 6975
ATCGATGTGGACGGCTCGAGAGTGTGGTTCACCGACTTGTGGAACTATTCCATTATCCCCTATCTCCTGGAAGCC 7050
GTCAGAGAAGGACTCCAGCTCTATGGAAGGCGCGCCCCCTGGGAGGATCCTGCCAAGTGGGTGATGGACACATAT 7125
CCATGGGCAGCCAGCCCACAACAGCACGAGTGGCCTCCCCTGCTGCAGTTACGGCCTGAGGATGTCGGCTTCGAC 7200
GGCTACTCCATGCCTCGGGAGGGATCGACAAGCAAGCAGATGCCCCCCAGTGATGCTGAAGGTGACCCGCTGATG 7275
AACATGCTGATGAGGCTGCAGGAGGCAGCCAACTACTCCAGCCCCCAGAGCTATGACAGCGACTCCAACAGCAAC 7350
AGCCATCACGATGACATCTTGGACTCCTCTTTGGAGTCCACTCTGTGACAGGGGCCCGGAGCCCAGCGCCCTCCT 7425
CTTCTCCTCACCGCATTCCACCTGCATCCCCCACATCACCCTGAAGATGACTTCCTGAGCCAGCCCCCAGCCACA 7500
GCCTTAGAGCTGCGGGAACACCGAGACCCCCCGTCCTTCAGCCTCGACCTGGGTGCAGGCATCCCGGGCCAGCTG 7575
CCTGCGGACCGCTTCCTTCCACAGCGAGAACTGCACTACCTTCTGTTGTACTTTAATTATTGTTTTGCCTTGTTG 7650
CTGTGACCTCCCTAAGACACTGAAGATACTTCTCGGGAAAGGATCATCGCCGTTGAAATGAAAAAAAAAAAAAAA 7725
AAAAAAAAAAAAAAAAAAAAAAAA 7748

At multiple positions heterozygous sequences have been observed. The ambiguities are denoted in the IUPAC IUB codes, which are as follows : R = A or G ; Y = C or T ; W = A or T ; M = C or A.

The region between position 5425 and 5433 is absent in cDNAs from Hela and colorectal adenocarcinoma tissue. Other cDNA sources are heterozygous (fragment present and absent) at this position.

cDNA from frontal cortex is heterozygous for the presence or absence of the region between 5924 and 6024. Absence of this fragment results in an out-of-frame deletion of 101 bp, resulting in a premature stop in translation.

The sequence in bold corresponds to the fragment in the 3'-UTR Hs-unc-53/2 that was used in RH mapping. The primers used to amplify SHGC-33456 are underlined.

Figure 1C (CONTINUED)

Three variants have been found for the 5' end of the gene. For these variants, the sequence from position 1 to position 366 should be replaced by one of the following sequences :

Variant 1

```
TGAAGAGGTGGTGCTGATTTCCTTGGCTGGCGGGAACTCTGTCTGGCTGTTGCATGCATCACTTTTGTGTGGGTT   75
ATTTTGTTCCTCTGTGGATTTGGAAGCATCGCTGAAGGAGAGAGAGGATTTTATTTCTGGGAAATGGAATCGGTT  150
TCTGAGTCCAGCCAACAGCAGAAGAGAAAGCCAGTTATCCACGGACTGGAAGATCAAAAGAGG               213
```

Variant 2

```
TGATACTTTGGGGTGCACATGGCTATTGATCTCTACTGCGGTTTGGCTTGTCTGTGGGGAATACATGAGCCCCGA   75
```

Variant 3

```
TAACAACTGGACTTTATTGAGTGTTTACCATGCACCAAGCCCTGGGCTAAACACTTCATCTGCAGGCTGTTCGTC   75
TTTACGGCAAACCCAGTAGGTAGGTATAACTATCCCCACTCTGCAGATGCAGAAACGGAGGCACAGAGTGTTTTG  150
GTAGCTAAACAAGCTCACCAGGAGGCTAGAAGGTGGCCACACCTAGCTGGCCCCCCTGACTCCACCAACTGCCTC  225
CCTTTGCTGTGTTGCATGCAAGAATGTGACTCCAAGTTTTTCCTTCCTTCTGGATCCAACTCTGGCTTCACTCTG  300
CTCAGCAACCAG                                                                  312
```

Figure 1d. Amino acid sequence of the protein encoded by the Hs-unc-53/2 gene

```
mPAILVASKmKSGLPKPVHSAAPILHVPPARAGPQPCYLKLGSKVEVSKTTYPSQIPLKSQVLQGLQEPAGEGLP   75
LRKSGSVENGFDTQIYTDWANHYLTKSGHKRLIKDLQQDVTDGVLLAQIIQVVANEKIEDINGCPKNRSQMIENI  150
DACLNFLAAKGINIQGLSAEEIRNGNLKAILGLFFSLSRYKQQQQQPQKQHLSSPLPPAVSQVAGAPSQCQAGTP  225
QQQVPVTPQAPCQPHQPAPHQQSKAQAEMQSRLSGPTARVSAAGSEAKTRGGSTTANNRRSQSFNNYDKSKPVTS  300
PPPPPSSHEKEPLASSASSHPGMSDNAPASLESGSSSTPTNCSTSSAIPQPGAATKPWRSKSLSVKHSATVSMLS  375
VKPPGPEAPRPTPEAMKPAPNNQKSMLEKLKLFNSKGGSKAGEGPGSRDTSCERLETLPSFEESEELEAASRMLT  450
TVGPASSSPKIALKGIAQRTFSRALTNKKSSLKGNEKEKEKQQREKDKEKSKDLAKRASVTERLDLKEEPKEDPS  525
GAAVPEMPKKSSKIASFIPKGGKLNSAKKEPMAPSHSGIPKPGMKSMPGKSPSAPAPSKEGERSRSGKLSSGLPQ  600
QKPQLDGRHSSSSSSLASSEGKGPGGTTLNHSISSQTVSGSVGTTQTTGSNTVSVQLPQPQQQYNHPNTATVAPF  675
LYRSQTDTEGNVTAESSSTGVSVEPSHFTKTGQPALEELTGEDPEARRLRTVKNIADLRQNLEETMSSLRGTQVT  750
HSTLETTFDTNVTTEMSGRSILSLTGRPTPLSWRLGQSSPRLQAGDAPSMGNGYPPRANASRFINTESGRYVYSA  825
PLRRQLASRGSSVCHVDVSDKAGDEMDLEGISMDAPGYMSDGDVLSKNIRTDDITSGYMTDGGLGLYTRRLNRLP  900
DGMAVVRETLQRNTSLGLGDADSWDDSSSVSSGISDTIDNLSTDDINTSSSISSYANTPASSRKNLDVQTDAEKH  975
SQVERNSLWSGDDVKKSDGGSDSGIKMEPGSKWRRNPSDVSDxSDKSTSGKKNPVISQTGSWRRGMTAQVGITMP 1050
RTKASAPAGALKTPGTGKTDDAKVSEKGRLSPKASQVKRSPSDAGRSSGDESKKPLPSSSRTPTANANSFGFKKQ 1125
SGSAAGLAMITASGVTVTSRSATLGKIPKSSALVSRSAGRKSSMDGAQNQDDGYLALSSRTNLQYRSLPRPSKSN 1200
SRNGAGNRSSTSSIDSNISSKSAGLPVPKLREPSKTALGSSLPGLVNQTDKEKGISSDNESVASCNSVKVNPAAQ 1275
PVSSPAQTSLQPGAKYPDVASPTLRRLFGGKPTKQVPIATAENMKNSVVISNPHATMTQQGNLDSPSGSGVLSSG 1350
SSSPLYSKNVDLNQSPLASSPSSAHSAPSNSLTWGTNASSSSAVSKDGLGFQSVSSLHTSCESIDISLSSGGVPS 1425
HNSSTGLIASSKDDSLTPFVRTNSVKTTLSESPLSSPAASPKFCRSTLPRKQDSDPHLDRNTLPKKGLRYTPTSQ 1500
LRTQEDAKEWLRSHSAGGLQDTAANSPFSSGSSVTSPSGTRFNFSQLASPTTVTQMSLSNPTMLRTHSLSNADGQ 1575
YDPYTDSRFRNSSMSLDEKSRTMSRSGSFRDGFEEVHGSSLSLVSSTSSVYSTPEEKCQSEIRKLRRELDASQEK 1650
VSALTTQLTANAHLVAAFEQSLGNMTIRLQSLTMTAEQKDSELNELRKTIELLKKQNAAAQAAINGVINTPELNC 1725
KGNGTAQSADLRIRRQHSSDSVSSINSATSHSSVGSNIESDSKKKKRKNWvneLRSSFKQAFGKKKSPKSASSHS 1800
DIEEMTDSSLPSSPKLPHNGSTGSTPLLRNSHSNSLISECMDSEAETVMQLRNELRDKEMKLTDIRLEALSSAHQ 1875
LDQLREAMNRMQSEIEKLKAENDRLKSESQGSGCSRAPSQVSISASPRQSMGLSQHSLNLTESTSLDMLLDDTGE 1950
CSARKEGGRHVKIVVSFQEEMKWKEDSRPHLFLIGCIGVSGKTKWDVLDGVVRRLFKEYIIHVDPVSQLGLNSDS 2025
VLGYSIGEIKRSNTSETPELLPCGYLVGENTTISVTVKGLAENSLDSLVFESLIPKPILQRYVSLLIEHRRIILS 2100
GPSGTGKTYLANRLSEYIVLREGRELTDGVIATFNVDHKSSKELRQYLSNLADQCNSENNAVDMPLVIILDNLHH 2175
VSSLGEIFNGLLNCKYHKCPYIIGTMNQATSSTPNLQLHHNFRWVLCANHTEPVKGFLGRFLRRKLMETEISGRV 2250
RNMELVKIIDWIPKVWHHLNRFLEAHSSSDVTIGPRLFLSCPIDVDGSRVWFTDLWNYSIIPYLLEAVREGLQLY 2325
GRRAPWEDPAKWVMDTYPWAASPQQHEWPPLLQLRPEDVGFDGYSMPREGSTSKQMPPSDAEGDPLMNMLMRLQE 2400
AANYSSPQSYDSDSNSNSHHDDILDSSLESTL                                              2432
```

Putative start methanionines at positions 1 and 10 are in lower cases. The residue at position 1018 (denoted by x) is encoded by an heterozygous sequence. Both residues Aspartic acid (D) or Glutamic acid (E) can be incorporated. The amino acid sequence VNE at position 1776 to 1778 is present or absent depending on the allele from which the protein is translated.

For translation of the 3 variants described in figure 1c, the aminosequence from position 1 to 89 has to be replaced by the following amino acid sequences :

Variant 1 25
mESVSESSQQQKRKPVIHGLEDQKR

Variant 2 19
mAIDLYCGLACLWGIHEPr

Variant 3 24
mQECDSKFFLPSGSNSGFTLLSNQ

Figure 1e. Nucleotide sequence of Hs-unc-53/3.

TAGAAGCATTTTCTTTGGCAGCAAGAAGATAATTTTATAGAAGCCATGCCTGTTCTTGGGGTTGCCTCAAAACTG 75
AGGCAGCCAGCTGTTGGGTCAAAGCCTGTGCATACTGCTCTTCCGATACCAAATCTTGGCACTACTGGGTCACAG 150
CACTGTTCTTCAAGACCTTTGGAACTTGCTGAAACAGAGAGCTCCATGCTTTCTTGTCAGCTTGCGTTAAAATCA 225
ACCTGTGAATTTGGAGAGAAGAAACCCCTCCAAGGAAAAGCCAAGGAGAAAGAAGACAGCAAGATTTACACTGAC 300
TGGGCCAACCACTACCTAGCAAAATCAGGCCACAAGCGGCTGATCAAGGACTTGCAACAAGACATTGCAGATGGA 375
GTACTCCTAGCAGAAATCATCCAGATTATTGCAAATGAAAAAGTTGAAGATATCAATGGATGTCCTAGAAGTCAG 450
TCTCAGATGATTGAAAATGTTGATGTCTGCCTTAGTTTTCTAGCAGCCAGAGGGGTAAATGTTCAAGGTCTATCT 525
GCTGAAGAAATAAGAAATGGAAACTTAAAAGCCATTCTAGGGCTGTTTTTCAGTTTATCTCGCTACAAGCAGCAA 600
CAACACCATCAACAACAGTACTATCAGTCCTTGGTGGAACTTCAGCAGCGAGTTACTCACGCTTCCCCTCCATCG 675
GAAGCCAGCCAGGCCAAAACCCAGCAAGATATGCAGTCCAGTCTGGCAGCCAGATATGCAACTCAGTCTAATCAC 750
AGTGGAATTGCAACCAGTCAAAAAAAGCCTACTAGGCTTCCAGGGCCCTCTAGGGTGCCTGCTGCAGGAAGCAGC 825
AGCAAGGTCCAGGGAGCCTCTAATTTAAATAGGAGAAGTCAGAGCTTTAACAGCATTGACAAAAACAAGCCTCCA 900
AATTATGCAAATGGAAACGAAAAAGATTCCTCCAAAGGACCTCAATCGTCTTCAGGTGTAAATGGTAACGTGCAG 975
CCTCCCAGTACTGCTGGGCAGCCTCCTGCCTCTGCCATCCCTTCTCCAAGTGCCAGCAAGCCCTGGCGCAGCAAG 1050
TCCATGAATGTCAAACACAGTGCCACCTCCACCATGTTGACTGTAAAGCAGTCAAGTACAGCCACCTCCCCCACA 1125
CCATCTTCAGACAGACTGAAGCCACCTGTCTCAGAAGGGGTCAAAACTGCTCCCTCAGGACAGAAATCCATGCTT 1200
GAGAAATTCAAGCTAGTCAATGCCCGGACTGCTTTACGCCCCCCGCAGCCTCCCAGTTCAGGACCTAGTGATGGT 1275
GGGAAGGATGATGATGCCTTTTCTGAATCTGGTGAAATGGAAGGTTTTAACAGTGGTCTGAATAGTGGTGGCTCA 1350
ACAAATAGCAGTCCCAAAGTGTCACCTAAGTTGGCCCCTCCAAAAGCTGGAAGCAAAAATCTCAGCAATAAAAAG 1425
TCTTTGCTACAGCCAAAGGAAAAAGAAGAAAAGAACAGGGACAAAAATAAAGTTTGCACTGAAAAACCAGTCAAA 1500
GAAGAGAAGGATCAGGTGACAGAGATGGCTCCAAAAAAGACCTCCAAAATTGCAAGCTTGATCCCTAAGGGCAGC 1575
AAGACAACAGCAGCTAAGAAGGAAAGCTTAATTCCGTCTTCCAGTGGTATTCCAAAACCAGGCTCTAAAGTTCCA 1650
ACAGTAAAGCAAACCATTTCACCTGGCAGCACAGCAAGCAAAGAGTCTGAGAAATTCAGGACTACCAAGGGGAGC 1725
CCTTCCCAGTCCTTATCTAAGCCTATAACCATGGAGAAAGCAAGTGCTTCTAGTTGTCCTGCCCCTTTGGAAGGA 1800
AGGGAAGCTGGCCAAGCTTCTCCTTCTGGTTCCTGTACCATGACAGTGGCACAAAGCAGTGGGCAGAGCACAGGA 1875
AATGGTGCTGTCCAACTCCCTCAACAGCAGCAACATAGCCACCCGAATACCGCGACAGTGGCACCATTCATTTAC 1950
AGGGCACATTCAGAAAATGAAGGTACCGCTTTACCATCGGCTGACTCCTGTACCAGTCCTACAAAGATGGACTTA 2025
TCATATAGTAAGACTGCTAAGCAGTGCCTGGAGGAGATATCTGGTGAAGGCCCTGAAACAAGAAGAATGAGAACA 2100
GTTAAAAACATAGCAGACTTGAGGCAGAATTTAGAAGAGACTATGTCCAGTCTTCGTGGGACTCAGATAAGCCAC 2175
AGCACCCTGGAGACAACATTTGACAGCACTGTGACAACAGAAGTTAATGGAAGGACCATACCCAACTTGACAAGT 2250
CGACCCACCCCCATGACCTGGAGGTTGGGCCAGGCATGTCCGCGACTTCAGGCGGGAGATGCTCCCTCCCTGGGT 2325
GCTGGCTATCCTCGCAGTGGTACCAGTCGATTCATCCACACAGACCCCTCGAGGTTCATGTATACCACGCCTCTC 2400
CGTCGAGCTGCTGTCTCTAGGCTGGGAAACATGTCACAGATTGACATGAGTGAGAAAGCAAGCAGTGACCTGGAC 2475
ATGTCTTCTGAGGTCGATGTGGGTGGATATATGAGTGATGGTGATATCCTTGGGAAAAGTCTCAGGACTGATGAC 2550
ATCAACAGTGGGTACATGACAGATGGAGGACTTAACCTATATACTAGAAGTCTGAACCGAATACCAGACACAGCA 2625
ACTTCCCGGGACATCATCCAGAGAGGGGTTCACGATGTGACAGTGGATGCAGACAGCTGGGATGACAGCAGTTCA 2700
GTGAGCAGTGGTCTCAGTGACACCCTTGATAACATCAGCACTGATGACCTGAACACCACATCCTCTGTCAGCTCT 2775
TACTCCAACATCACCGTCCCCTCTAGGAAGAATACTCAGCTGAGGACAGATTCAGAGAAACGCTCCACCACAGAC 2850
GAGACCTGGGATAGTCCTGAGGAACTGAAAAAACCAGAAGAAGATTTTGACAGCCATGGGGATGCTGGTGGCAAG 2925
TGGAAGACTGTGTCCTCTGGACTTCCTGAAGACCCCGAGAAGGCAGGGCAGAAAGCTTCCCTGTCTGTTTCACAG 3000
ACAGGTTCCTGGAGAAGAGGCATGTCTGCCCAAGGAGGGGCGCCATCTAGGCAGAAAGCTGGAACAAGTGCACTC 3075
AAAACACCCGGGAAAACCGATGATGCCAAAGCTTCTGAGAAAGGAAAAGCTCCCCTAAAAGGATCATCTCTACAA 3150
AGATCTCCTTCAGATGCAGGAAAAAGCAGTGGAGATGAAGGGAAAAAGCCCCCCTCAGGCATTGGAAGATCGACT 3225
GCCACCAGCTCCTTTGGCTTTAAGAAACCAAGTGGAGTAGGGTCATCTGCCATGATCACCAGCAGTGGAGCAACC 3300
ATAACAAGTGGCTCTGCAACACTGGGTAAAATTCCAAAATCTGCTGCCATTGGCGGGAAGTCAAATGCAGGGAGA 3375
AAAACCAGTTTGGACGGTTCACAGAATCAGGATGATGTTGTGCTGCATGTTAGCTCAAAGACTACCCTACAATAT 3450
CGCAGCTTGCCCCGCCCTTCAAAATCCAGCACCAGTGGCATTCCTGGCCGAGGAGGCCACAGATCCAGTACCAGC 3525
AGTATTGATTCCAACGTCAGCAGCAAGTCTGCTGGGGCCACCACCTCGAAACTGAGAGAACCAACTAAAATTGGG 3600
TCAGGGCGCTCGAGTCCTGTCACCGTCAACCAAACAGACAAGGAAAAGGAAAAAGTAGCAGTCTCAGATTCAGAA 3675
AGTGTTTCTTTGTCAGGTTCCCCCAAATCCAGCCCCACCTCTGCCAGCGCCTGTGGTGCACAAGGTCTCAGGCAG 3750
CCAGGATCCAAGTATCCAGATATTGCCTCACCCACATTTCGAAGgttgtttggtgccaaggcaggtggcaaatct 3825
gcctctgcacctaatactgagggtgtgaaatcttcctcagtaatgcccagccctagtaccacattagcgcggcaa 3900
ggcagtctggagtcaccgtcgtccggtacgggcagcatgggcagtgctggtgggctaagcggcagcagcagccct 3975
ctcttcaataaaccctcagacttaactacagatgttataagcttaagtcactcgttggcctccagcccagcatcg 4050
gttcactctttcacatcaggtggtctcgtgtgggctgccaatatgagcagttcctctgcaggcagcaaggatact 4125
ccgagctaccagtccatgactagcctccacacgagctctgagtccattgacctcccctcagccatcatggctcc 4200
ttgtctggactgaccacaggcactcacgaggtccagagcctgctcatgagaacgggtagtgtgagatctactctc 4275
tcagaaagcatgcagcttgacagaaatacactacccaaaaagggactaagATATACCCCATCATCTCGGCAGGCC 4350
AACCAAGAAGAGGGCAAAGAGTGGTTGCGTTCTCATTCTACTGGAGGGCTTCAGGACACTGGCAACCAGTCACCT 4425
CTGGTTTCCCCTTCTGCCATGTCATCTTCTGCAGCTGGAAAATACCACTTTTCTAACTTGGTGAGCCCAACAAAT 4500
TTGTCTCARTTTAACCTTCCCGGGCCCAGCATGATGCGCTCAAACAGCATCCCAGCCCAAGACTCTTCCTTCGAT 4575
CTCTATGATGACTCCCAGCTTTGTGGGAGTGCCACTTCTCTGGAGGAAAGACCTCGTGCCATCAGTCATTCGGGC 4650
TCATTCAGAGACAGCATGGAAGAAGTTCATGGCTCTTCATTATCACTGGTGTCCAGCACTTCTTCTCTTTACTCT 4725
ACAGCTGAAGAAAAGGCTCATTCAGAGCAAATCCATAAACTGCGGAGAGAGCTGGTTGCATCACAAGAAAAAGTT 4800
GCTACCCTCACATCTCAGCTTTCAGCAAATGCTCACCTTGTAGCAGCTTTTGAAAAGAGCTTAGGGAATATGACT 4875

Figure 1e (CONTINUED) 9/56

GGCCGATTGCAAAGTCTAACTATGACAGCGGAACAAAAGGAATCTGAACTTATAGAACTAAGAGAAACCATTGAA 4950
ATGCTGAAGGCTCAGAATTCTGCTGCCCAGGCGGCTATTCAGGGAGCACTGAATGGTCCAGACCATCCTCCCAAA 5025
GATCTTCGCATCAGAAGACAGCATTCCTCTGAAAGTGTTTCTAGTATCAACAGTGCCACAAGCCATTCCAGTATT 5100
GGCAGTGGTAATGATGCCGACTCCAAGAAGAAGAAAAAGAAAAACTGGGTGAactctagaggaagtgagctgaGA 5175
AGTTCTTTCAAACAAGCCTTTGGGAAGAAAAAGTCCACCAAGCCTCCTTCATCACATTCTGACATTGAAGAGCTT 5250
ACTGATTCATCCCTTCCGGCATCCCCCAAGTTACCCCATAATGCTGGTGACTGTGGCTCAGCATCCATGAAGCCC 5325
TCACAATCTGCTTCAGCgtcacccctttgtctggccaccaaagaaacgacaaaatggccctgtgatctacaagcat 5400
agatctcgGATCTGTGAATGCACAGAAGCTGAGGCAGAGATAATTCTGCAGCTGAAGAGCGAGCTCAGAGAAAAG 5475
GAATTAAAATTAACGGATATTCGGCTGGAGGCCCTCAGCTCTGCTCATCATCTTGATCAGATCCGGGAAGCCATG 5550
AACCGGATGCAGAATGAAATTGAAATACTGAAAGCTGAAAATGACCGGTTGAAGGCAGAAACTGGTAACACAGCT 5625
AAGCCTACTCGGCCACCGTCAGAATCCTCAAGCAGCACCTCCTCTTCATCTTCCAGGCAGTCATTAGGACTTTCT 5700
CTAAACAATTTGAACATCACAGAGGCTGTTAGCTCAGATATTTTGCTAGATGATGCTGGTGATGCAACTGGACAT 5775
AAAGATGGCCGCAGTGTGAAAATTATAGTCTCCATAAGCAAGGGCTATGGTCGAGCAAAGGACCAAAAATCTCAG 5850
GCATATTTGATAGGATCCATTGGTGTTAGTGGAAAAACCAAGTGGGATGTCTTAGATGGTGTAATAAGACGTCTC 5925
TTTAAGGAATATGTATTCCGAATTGATACATCCACTAGCCTTGGTCTGAGCTCTGACTGCATTGCTAGCTACTGT 6000
ATAGGAGACTTAATTAGATCCCATAACCTAGAAGTGCCTGAATTGCTGCCTTGTGGATACCTTGTTGGAGATAAT 6075
AACATCATCACTGTGAACCTCAAAGGGGTAGAAGAAAATAGTTTGGACAGTTTTGTTTTTGATACGCTGATTCCT 6150
AAACCAATTACCCAAAGGTACTTTAACTTGTTGATGGAGCATCACAGAATTATACTCTCAGGACCGAGTGGTACT 6225
GGAAAGACCTATTTGGCAAACAAACTTGCTGAATATGTAATAACCAAATCTGGAAGGAAAAAAACAGAGGATGCA 6300
ATTGCCACTTTTAATGTGGACCACAAGTCAAGTAAGGAATTGCAACAATATCTAGCTAACCTGGCTGAACAGTGC 6375
AGTGCTGATAATAATGGAGTGGAGCTCCCAGTTGTAATAATTCTTGATAATCTTCATCATGTGGGCTCTCTGAGT 6450
GATATCTTCAATGGTTTTCTCAATTGTAAATACAACAAATGTCCATATATTATTGGAACAATGAATCAGGGAGTT 6525
TCTTCATCACCAAATCTAGAGCTGCATCACAATTTCAGGTGGGTATTATGTGCAAATCATACAGAACCAGTGAAA 6600
GGCTTTTTAGGCAGATATCTTCGAAGAAAACTCATAGAGATAGAAATTGAAAGGAACATTCGCAATAATGACCTA 6675
GTCAAAATTATAGATTGGATTCCGAAGACGTGGCATCATCTCAACAGTTTTTTGGAAACACACAGTTCTTCTGAC 6750
GTTACCATTGGTCCCCGACTATTCCTTCCTTGCCCCATGGATGTAGAAGGTTCTAGAGTATGGTTCATGGATCTC 6825
TGGAACTATTCTTTAGTACCTTATATTCTGGAGGCAGTGAGAGAGGGTCTTCAGATGTATGGGAAACGCACACCA 6900
TGGGAAGATCCTTCAAAGTGGGTGCTTGACACATATCCATGGAGCTCAGCAACTCTGCCTCAGGAGAGCCCAGCC 6975
TTACTTCAGCTGCGACCAGAAGATGTTGGGTATGAAAGCTGCACATCCACTAAGGAAGCCACAACCTCAAAGCAC 7050
AGCACACAAAGCTGCGACAGCGAAAGCACCAGCCACCATGAAGACATTTTGGATTCATCTCTTGAATCTACCCTC 7200
TAGAGGGTGAAAAAAGTTAAGGGAAAAGACTTTGCTTTTAAAAAAATGTTTCAAAAGAAAGGTATTTTCACTAAA 7275
CCACTGCCAGTATAAAAGCACCCTGTCAAGGGCCCTGACCCAGAGTTGTGGTCTCCAAGGAGGCAGCAGAACTAA 7350
GTCTGAACCGCCAAGATGCTAAATTGCAATGGAAGCTTAACTTTAGTTTATTTCTAAGCATTTTTTATATCTGTG 7425
GAGTAATAGAAAGCTCCATTACTCAACTGGAAAGGACCCTAATGACAGGGCAACTGAACAGATTGCACATGGGAT 7500
AGCCAAACTGGACTTTCTTTGTTTCCTCTTTAAAAGTTTACAATGCAGACCATTTTTTGTCCCTTCCTTTTGTTT 7575
CCTCTGAGGGGCTGTTCGCCCCAGGCAGGGTCCATCTTTCTGATCTGTCCAACCTCCTTTGTGCCACACGGTGCT 7650
GGTCACAGGGCTTCAGTAGTGTTTGTGTTGTGCGCTCACCCCATTCCAGAACAAATCCAAGAGGCCAGTCCTCCA 7725
TAAGCACAAATGGAATTGTGCAACCACCAGAAAAACACTACTGTGGCAAACTGGAGAAGTGCCAATTTAATTCTA 7800
ACTGCCACGTTCTCATGATGTGCTCCACCAACTTTTTAGTATATGAGTCACTGGTTTTATAAGGTTGTTTTTACC 7875
ACAGTGGTCTTTTTAAACCACCTGCCCACTCCCTTAACAAGAGTTTTATACCAATTATTAGTCAACACTGATAAA 7950
AGGCTTTTTTAGGGCTTTATTTGTTTGAGCCTTTTCAGTGAAAGAAGGAACATTTCCTATGGTGCTGTCTCACTG 8025
CCTTAAAACAGATTTCTATGACAGTTTAACAGTTGGTTTAAATCCTAAACCATTGGTAATTTCCACTGTCTTTTC 8100
ATTTACAACCAAGCAACACCAGTTAACATAGTAGCCTCATCTCTATATATCTTTCTCTTTTTTTTTTTTTTTGAAG 8175
AAATGGATAGGAGAAAGATCAGTATTTTTAGCCTTGTGAATAGATCGCTTTGCCTATCCTCCAAAATATTAAAAT 8250
AACCCAGAAATGCTCTTTGACCGTCACTTAAAACCTAAGACATGTGGCGAAATTCCATCCAGTTCTAAGTGAAAG 8325
AGTTTCAGAAGGCAGGAGATTTTGAATTATTATCCAGCAGGGCTGGAAGCACTAGATGCAGCATGAGCACAACTA 8400
TTCGGCTTTCCTTCCCTATTGTTTTTGTTTTTTTAATGAGTTTTGACGCATGTTGTTTTGATTGCTATTGTTGTA 8475
CATGAGAAATTCAGCATTAAAGAACACTGAAGCGGTAAGGTCACTGTGGAAGAGGAAGCGTTTATACTGTAAAAG 8550
AAGGTTAGATTTGCACAGTCTACTGGGTAGGTATTGTAAATAATAATTTTTAAAACTTGCACAAATCAAAACAAA 8625
CACAAACAAAATTGTATTTTATCCTGTTGGTGTTAAGAGGTGTTTCACTTGCTGAGATTTCCTGTACATTGCAAA 8700
CAAATACAGAATGCAAACCCTCAAAGCTGTATTATCTGGTGTGTTTGTCCTGTATTTACAGTTGTTTTTGACTAT 8775
GCAGGAGCTATCAGTGCTAGAGTGAGCATGCTTCAAAACTGTACATGAAGCCAATATATTTTTGGATAAGTAAAA 8850
AAAAAAAAAAAAAACTCGAGGGGGGGCCCGGTACCCAATTCGCCCTATAGTGAGTCGTATTACAATTCACTGGCC 8925
GTCGTTTTACAACGTCGTGACTGG 8949

The region from position 3795 to 4325 consists of two blocks (3795 to 4283 and from 4284 to 4325) that independently can be present or absent in cDNA molecules from frontal cortex tissue. Frontal cortex is also heterozygous for the region from 5153 to 5173. The region from 5343 to 5408 is absent in frontal cortex, but heterozygously present in hart cDNAs.
The nucleotide sequence in heterozygous at position 4509. R is the IUB IUPAC code for A or G. Amino acid sequence is not affected.
An alternative 5′ end has been observed. In this variant the sequence from position 1 to 288 is replaced by the following DNA sequence :
TAGTTTGCTGCTTTTTTGAAGAGATTCCATTTTGAAGGGCAAGAACCTAATGTGATGGATTTATCTTCAGAAATG 75
AACAGACATGGGAAGAATCCAGTGAGTCACAAGCTAGAAGATCAGAAGAAG

Figure 1f. Protein sequence encoded by the Hs-unc-53/3 gene

```
mPVLGVASKLRQPAVGSKPVHTALPIPNLGTTGSQHCSSRPLELAETESSmLSCQLALKSTCEFGEKKPLQGKAK   75
EKEDSKIYTDWANHYLAKSGHKRLIKDLQQDIADGVLLAEIIQIIANEKVEDINGCPRSQSQMIENVDVCLSFLA  150
ARGVNVQGLSAEEIRNGNLKAILGLFFSLSRYKQQQHHQQQYYQSLVELQQRVTHASPPSEASQAKTQQDMQSSL  225
AARYATQSNHSGIATSQKKPTRLPGPSRVPAAGSSSKVQGASNLNRRSQSFNSIDKNKPPNYANGNEKDSSKGPQ  300
SSSGVNGNVQPPSTAGQPPASAIPSPSASKPWRSKSMNVKHSATSTMLTVKQSSTATSPTPSSDRLKPPVSEGVK  375
TAPSGQKSMLEKFKLVNARTALRPPQPPSSGPSDGGKDDDAFSESGEMEGFNSGLNSGGSTNSSPKVSPKLAPPK  450
AGSKNLSNKKSLLQPKEKEEKNRDKNKVCTEKPVKEEKDQVTEMAPKKTSKIASLIPKGSKTTAAKKESLIPSSS  525
GIPKPGSKVPTVKQTISPGSTASKESEKFRTTKGSPSQSLSKPITMEKASASSCPAPLEGREAGQASPSGSCTMT  600
VAQSSGQSTGNGAVQLPQQQQHSHPNTATVAPFIYRAHSENEGTALPSADSCTSPTKMDLSYSKTAKQCLEEISG  675
EGPETRRMRTVKNIADLRQNLEETMSSLRGTQISHSTLETTFDSTVTTEVNGRTIPNLTSRPTPMTWRLGQACPR  750
LQAGDAPSLGAGYPRSGTSRFIHTDPSRFMYTTPLRRAAVSRLGNMSQIDMSEKASSDLDMSSEVDVGGYMSDGD  825
ILGKSLRTDDINSGYMTDGGLNLYTRSLNRIPDTATSRDIIQRGVHDVTVDADSWDDSSSVSSGLSDTLDNISTD  900
DLNTTSSVSSYSNITVPSRKNTQLRTDSEKRSTTDETWDSPEELKKPEEDFDSHGDAGGKWKTVSSGLPEDPEKA  975
GQKASLSVSQTGSWRRGMSAQGGAPSRQKAGTSALKTPGKTDDAKASEKGKAPLKGSSLQRSPSDAGKSSGDEGK 1050
KPPSGIGRSTATSSFGFKKPSGVGSSAMITSSGATITSGSATLGKIPKSAAIGGKSNAGRKTSLDGSQNQDDVVL 1125
HVSSKTTLQYRSLPRPSKSSTSGIPGRGGHRSSTSSIDSNVSSKSAGATTSKLREPTKIGSGRSSPVTVNQTDKE 1200
KEKVAVSDSESVSLSGSPKSSPTSASACGAQGLRQPGSKYPDIASPTFRRlfgakaggksasapntegvkssvm 1275
pspsttlarqgslespssgtgsmgsagglsgsssplfnkpsdlttdvislshslasspasvhsftsgglvwaanm 1350
ssssagskdtpsyqsmtslhtssesidlplshhgslsglttgthevqsllmrtgsvrstlsesmqldrntlpkkg 1425
LrYTPSSRQANQEEGKEWLRSHSTGGLQDTGNQSPLVSPSAMSSSAAGKYHFSNLVSPTNLSQFNLPGPSMMRSN 1500
SIPAQDSSFDLYDDSQLCGSATSLEERPRAISHSGSFRDSMEEVHGSSLSLVSSTSSLYSTAEEKAHSEQIHKLR 1575
RELVASQEKVATLTSQLSANAHLVAAFEKSLGNMTGRLQSLTMTAEQKESELIELRETIEMLKAQNSAAQAAIQG 1650
ALNGPDHPPKDLRIRRQHSSESVSSINSATSHSSIGSGNDADSKKKKKKNWVnsrgselRSSFKQAFGKKKSTKP 1725
PSSHSDIEELTDSSLPASPKLPHNAGDCGSASMKPSQSASAsplvwppkkrqngpviykhrsrICECTEAEAEII 1800
LQLKSELREKELKLTDIRLEALSSAHHLDQIREAMNRMQNEIEILKAENDRLKAETGNTAKPTRPPSESSSSTSS 1875
SSSRQSLGLSLNNLNITEAVSSDILLDDAGDATGHKDGRSVKIIVSISKGYGRAKDQKSQAYLIGSIGVSGKTKW 1950
DVLDGVIRRLFKEYVFRIDTSTSLGLSSDCIASYCIGDLIRSHNLEVPELLPCGYLVGDNNIITVNLKGVEENSL 2025
DSFVFDTLIPKPITQRYFNLLMEHHRIILSGPSGTGKTYLANKLAEYVITKSGRKKTEDAIATFNVDHKSSKELQ 2100
QYLANLAEQCSADNNGVELPVVIILDNLHHVGSLSDIFNGFLNCKYNKCPYIIGTMNQGVSSSPNLELHHNFRWV 2175
LCANHTEPVKGFLGRYLRRKLIEIEIERNIRNNDLVKIIDWIPKTWHHLNSFLETHSSSDVTIGPRLFLPCPMDV 2250
EGSRVWFMDLWNYSLVPYILEAVREGLQMYGKRTPWEDPSKWVLDTYPWSSATLPQESPALLQLRPEDVGYESCT 2325
STKEATTSKHIPQTDTEGDPLMNMLMKLQEAANYSSTQSCDSESTSHHEDILDSSLESTL                2385
```

Regions corresponding to heterozygous sequences encoding presence or absence of this region are in lower case letters. These regions are from 1326 to 1413 ; from 1414 to 1427 ; from 1703 to 1709 and from 1768 to 1788.

Putative start methionines at positions 1 and 51 are indicated in lower case.

For the variant mentioned in figure 1e, the amino acid sequence from position 1 to 81 has to be replaced by the following amino acid sequence :

```
mDLSSEmNRHGKNPVSHKLEDQKX                                                     24
```

Figure 1g. Nucleotide sequence of a 4984 bp fragment from BAC 585E09 (contains part of the genomic sequence of Hs-unc-53/1) extending the sequence derived from cDNA libraries shown in figure 1a.

```
TTCCTGATCTCAAGAGTTACTCCTTCCCCTACAAAGCCCTCAGCCCCCTCCCCAGTCAACGCTAGGCCCCTTCTC    75
TCCAAGCCACCCGTGTCCTACCCCCATCCCCTACCTCCTGGGCTCAGGAGGGCAACCTTGAGCCTCAGAGACTGA   150
AGTAGGGTGGGACTGGGAGTTTCCTGGGGGAAAACCAAAGACGGTTTGGGGTTGGGGGAGGGGAATGAGCACCCT   225
GGATACCATTCCTCCACCCCTCTCCCGACATCTCTCTCAGGCCCACGGGCCCACTTTCCCTCCCGCATTCTGAGC   300
CGCCCTCCCTCCGTCTCTTTTACCTGCACCTCCACACCTCCTCAACAGATCTTTATCCTGGACACGGCAGGGGGT   375
CCCCGTGCCCTCCGAGAATCCAAGAACCCGCCCCGCTTCTACGCGGAAAGCTGGGAGAAAAACTGCTTTTCCTTT   450
ATTTCCCCCTACCCCCCACTCATCCGCCCCTGGAGCTCCGCTCGCAGATACCTCCCCCTCCCGAGCCAGAAATAG   525
ACACACTATCTCTCCCCCACCTCCCTCCCCGTGCGCACACTCGCTCCCCTCCTCCTGTTTGCTCCCCGCCTTCCC   600
CTTCCCTCCTTCCTCTGCTCGGAGCTGCAGCCTGCAGCCTCGACTCGGGCTGGCTGGCTGGCTGAGTGCGGCCGG   675
GGCGCTGCCCGGCAGTGCGGTGTCCACGGGACTGACAGGCAGGCAGGCAGGCCGCGGGCTGGGATCCGGACACCA   750
AAGCAAAAGCACCGCTGGGCGCCGGAGGAGCCGCGGGGCTTCCATCCTTCCTTTGACTGATTTTTAAATTTTAAT   825
TTGTATTTTCCCCGCCGCCCCGCCCCTTTTCCTCCGACCCCGCCCTATCGCTCCCCGGCTTCCCTGCTCTTTCCT   900
TTTTCCCGGCTTCCTTCCTCGCGTTTCTTTCCCCTGCGCCCTCGGCTTGCCTCTCTCCCTCCTCCCTCGCTCTCT   975
CCCCCTTCTCTCCCCTTCTTCCTCGGTTTCTTCCGTCCTCTCTCTCCCCCTCCTCCTCCCCCGCCTCCTCCTCCT  1050
GCGCTCCCGCCCCCTGCCCCCTCCCCCCGTGCCTGCAGACGCGCGGATCGTCCATGCGCTCCTCGCGGGCAGAAT  1125
GCTGGGCAGCAGCGTCAAGAGCGTGCAGCCCGAGGTGGAGCTGAGCAGCGGCGGCGGCGACGAGGGCGCGGACGA  1200
ACCGCGGGGCGCCGGCAGGAAGGCGGCAGCGGCGGACGGCAGAGGCATGCTGCCCAAGCGCGCCAAGGCGCCCGG  1275
CGGCGGCGGCGGCATGGCCAAGGCCAGCGCGGCTGAGCTGAAGGTCTTCAAGTCCGGCAGCGTGGACAGCCGTGT  1350
CCCCGGCGGGCCGCCCGCCTCCAACCTGCGCAAGCAGAAGTCACTCACCAACCTCTCTTTTCTCACGGACTCCGA  1425
GAAAAAGCTGCAGCTTTATGAGCCCGAATGGAGCGACGATATGGCCAAGGCGCCCAAAGGCTTAGGCAAGGTGGG  1500
GTCCAAGGGCCGTGAAGCTCCGCTGATGTCCAAGACGCTGTCCAAGTCGGAGCACTCGCTCTTCCAGGCCAAGGG  1575
CAGCCCGGCGGGCGGTGCCAAGACCCCCCTGGCTCCGCTCGCGCCCAACCTGGGAAAGCCGAGCCGGATCCCTCG  1650
AGGACCCTATGCGGAGGTCAAGCCGCTCAGCAAGGCGCCTGAAGCGGCCGTGAGCGAAGATGGCAAATCGGACGA  1725
CGAGCTGCTCTCCAGCAAGGCCAAGGCGCAAAAGAGCTCTGGGCCTGTCCCCTCTGCCAAGGGCCAGGAGGAGCG  1800
CGCCTTCCTCAAGGTGGACCCCGAGCTGGTGGTGACCGTGCTGGGAGACCTGGAGCAGCTGCTCTTCAGCCAGAT  1875
GCTGGGTAAGTCCTGCCGCCCCGCCCCGCCCCGCCCCTGGCTTTCTCCTAACCAGCTGCTGGGGAAGGTGTGGGG  1950
AAAGCGAAGCCCCCTCCCCTTGCGCCTTCCCGGAGGGCCCTCCTGTTCACGATCAGGCTGTGATGGGCATTGCGC  2025
CCAGATGTGCTGAGCTGGCCCACCTCCAGATGCGCATGGCTCAAGTGTACCTTCTTAAAGACATTACAGGCGGGA  2100
ACCCGGGCTCGACTCTGGCCTGCCCGAGGTGAGGCTGGACAATGGGATGGGGGGTGAGGGGGTTACAGGCTCTCA  2175
GAAATAGAGCCAGAATCCCAATATGGCAAAACCTGGGACTGGTGGAAACCTCCGTTGTGGTGTGGCCTGCGCTTG  2250
ACAGGAGCATCCCGCATTGCAAGGGGAGCGTCCAGCGAGAGCCCGGATCTAGAGGACAGATGTGGGAGAGCAGAT  2325
GTGAGGGCTGATTGGCCCCGGAACACAGCTGAGGCTCCACTTCCTCTGTGGATCCCGAGTGGGAGCGCAAGTCGG  2400
ATTTCCCCGCGGTGTGAGGATTCTGGCTAAAAGAAGCGTCTAGGGCCGGGGGCGGGCGGGCTGCCAGCTGTGCGC  2475
ATCTGGGCGCATGTCCGATACCTCAGCCCCGGCTCTGGCCCCAACCCCTACACCCGCAGGTCTTTTAGGGCGTGT  2550
CGAAAGCTCTGGGCGTTAGCGCCGAGACTCCTGTTTGACCGGGAAGCCTTTGCCCTGTGGCTAATGGAAGCCGAG  2625
CAGGCGGGAAGGGAGGAACAAAGCTTGCTCGAGTGGAGGAAGCGCGCAGAGCTGTTCCATTGTTCTCCGTGCCTG  2700
AAGAGTCTATGCAAAAAAACCCGAAGCGGGCCCCGGAACTGCTCTTTCTCTCCCCGGAGAGCCCCTGCCCTCAGA  2775
GAGGAATAGATCTGGGATGTGCCGGACGCCAGGCGGCCATGCCTCGGGAACTGGCACGGGCCCTCTTGGGGCCAC  2850
GGAACAAGGACGGTGGGGCCTGGTGCCCAGGCGAGCTGCTTTGGCTGCGCGGACTTGTTGCGGTGGCTGGGTTGT  2925
GGGTCCTCCGGCGCCGAGGGACCCGAGCTTCCTGGGTACCCGGCAGGCTGCCCGCCCGCTGGGGGCTGGGAAGGG  3000
GCGTGCCAATGCGCGTGTGAAAGGGCGGGGCCGAGTGACGGGCTTGGTGGGTGGGGAACATGCAAGAGCTCGCCG  3075
GGCGGCCCTGGAGAATGCGAAGCCGGAGGAGACCGGTTCGGCCTGCTGCAGTCTTCCTAGGAACCCTCGACTCCT  3150
GTGTGGGCTAGGATGAGGGTCCTCTGACAGGGGCAAGGATTTGGGCCTTTGGAGAACCGATCCTTACGCAGGAGG  3225
CCGCAAATGGGCTTTGCAGGGGCAATCAGGAGACTGGACAAGGGCAAAAGAAGAGCAGCCTTTTCCCCTGGGAGC  3300
CCCTCCTGAAGGTGGGGATGGCTGGGTGGGTGCGGAAGCTGACCAGGCAGCCTCACTCTGCAAAGGGAATGTGCC  3375
ACCCGGTCCTCAGTGTGGGGCTGAGCCTGTCAAAGGCCCTGCCTCAGTGAATGGGGCAAGAGAGACAATAAGGGA  3450
AAAAATTAATAAATTTTTGGCAGGCACCATGGCAGGCACCAAGGAGGGATATGGACAAAATGCAACTGGCCCATG  3525
TGATAGAGAGCCCTGCTGAGGGACTGAAAGCAGGGTGAGAGAGGAAGGGACCGTGTGTGTGTGTGTGTGTGTGTG  3600
TGTGTGTGTGTGTGTGTGTGTGTGTGTCAGACTGAGAGAGCGACTTGGCGAGGGAGGGCAAGGGAGTATCGGG  3675
GCACAGAATAGCAGAAGGCACAGAACCCTTTTCAGGGTCAAGGCTTTACTTGTTGGGGATAACTTAGCTGGTCTG  3750
GGTCCTCTCCAGACCTGGATGCCCTCACACTGTCCCAGAAGCTGACTGCCCATTGAAGCCCTCTTAGTTGCTCCT  3825
CAAGAGGAGCCAAACAGGTCTGAGCTGGCTAGGGAGATGGGAGGAGGGGCAGGAGTGGGGAGGAGGGCAGGTGCA  3900
GGGAGGGCGAGAGAGGAGGAGAAGCTGAGCTGTGGTCCCTTATTCCTGCTTAGCAGTTGTCACTTCTCAAAGCAC  3975
ACTGACACTTTCAGTAACCTCGGAAGTGAGGAGAGAACACCTCCACTTCCCAGTTGGGGAAATGCAGAGTCAAAA  4050
GCATTGAGGGCCTTAAAGGCATCTATGAGTTCATGGTGGAAGGGAGATTCCACATTGATCTCCTGAGGACTAAGT  4125
CGCAGCTCTTCCTAGGAGACCTGATTGAGAGAGGAAGAGTCAGCAGGCAGAGGGACCTGCCCAAGGCTATGTCTA  4200
CTGGGTATGGGCCACCAGAACTGCCTCGATGACCCTACAGAGGGCTGAGGGGCTTAGCTCTCTGGGGTGGGGAGA  4275
GAAGGGTGGAAACTCCCAAATCTGCCTGTCTCCAGCTGAGAGGACCCAAAGTTGGGGGGTGGGGAGTTGGTTCAG  4350
GCTGTAGCAAGGCAGAGCCTGGTGTCAAACAGTGGTAGGGAGGAAAGGAGGGGAGTTGGTGACCTCCAAACTAAG  4425
CTTTTCCCTGTGTGAAGGGCAGAGGGTAGACTGCCTGGGGGAGGGGTAGAGGGAGAGGAACTACAGAGGGAATTC  4500
GTCTTCCAGAGCCAATGATGGTGGTGTTCAGGTATCAGACAGGCCCTCAGTGTACAGCAGGGTGGCCTCTGGGGA  4575
GAAGAATGGTGACTTGATGTTTCAGGATTGTGATTGAAGACACTGGGCATTTGTCCCCACCTCAGTGGGGCTCAG  4650
```

Figure 1g (CONTINUED)

```
TGTCCAGTTATGTTCACTCCATAGTACCATCCTAGATCCAAGAGGCTGCCAAGAATCAATTTCTGAGGCGGAGGG 4725
AGGGGGTGGGAGTGAGGCAGCTTCAAGTCAGAGCCTTTCTGTAATAAGAGGGAAGGACTGAAACCTGATCATCCC 4800
CTTCCCAGAAATCAGCTGGGGTCCCAGATGGTCTAGGCAGGCTCCCTGTCCCTTCGCTAACCTTGGAAGCTGCCA 4875
AATAACTAGGGCCCCACTGGGGAACCCTAGCAACTTGGAAGACTGAGGAGTGAGTACCGAGGGCAAATGGGCTAA 4950
TTCCAGGAATTAGATGCCTCTGGACCCTGGCCCG                                           4984
```

The sequence shown in figure 1a starts at position 1246. Upstream in the same reading frame as used for the translation of the DNA sequence in fig 1a into the protein sequence of fig 1b, a stop codon is found at position 815. A first putative start codon (ATG) can be found at position 1124. Assuming this start codon, the protein sequence from fig 1b is extended by the sequence
MLGSSVKSVQPEVELSSGGGDEGADEPRGAGRKAAAADGRG

Intronic sequence has been found to start at position 1881.

Figure 1h. Illustration of a 5'-deletion variant of Hs-unc-53/3 discovered by Nagase et al., (1999, DNA Res. 6:63-70).

>KIAA0938 protein, amino acid sequence

MCVTKKLFFIVQRTIFVGCVIWKFCLHYVLRGFLCFNSMQLDRNT
LPKKGLRYTPSSRQANQEEGKEWLRSHSTGGLQDTGNQSPLVSPSAMSSSAAGKYHFSN
LVSPTNLSQFNLPGPSMMRSNSIPAQDSSFDLYDDSQLCGSATSLEERPRAISHSGSFR
DSMEEVHGSSLSLVSSTSSLYSTAEEKAHSEQIHKLRRELVASQEKVATLTSQLSANAH
LVAAFEKSLGNMTGRLQSLTMTAEQKESELIELRETIEMLKAQNSAAQAAIQGALNGPD
HPPKDLRIRRQHSSESVSSINSATSHSSIGSGNDADSKKKKKKNWVNSRGSELRSSFKQ
AFGKKKSTKPPSSHSDIEELTDSSLPASPKLPHNAGDCGSASMKPSQSASAICECTEAE
AEIILQLKSELREKELKLTDIRLEALSSAHHLDQIREAMNRMQNEIEILKAENDRLKAE
TGNTAKPTRPPSESSSSTSSSSSRQSLGLSLNNLNITEAVSSDILLDDAGDATGHKDGR
SVKIIVSISKGYGRAKDQKSQAYLIGSIGVSGKTKWDVLDGVIRRLFKEYVFRIDTSTS
LGLSSDCIASYCIGDLIRSHNLEVPELLPCGYLVGDNNIITVNLKGVEENSLDSFVFDT
LIPKPITQRYFNLLMEHHRIILSGPSGTGKTYLANKLAEYVITKSGRKKTEDAIATFNV
DHKSSKELQQYLANLAEQCSADNNGVELPVVIILDNLHHVGSLSDIFNGFLNCKYNKCP
YIIGTMNQGVSSSPNLELHHNFRWVLCANHTEPVKGFLGRYLRRKLIEIEIERNIRNND
LVKIIDWIPKTWHHLNSFLETHSSSDVTIGPRLFLPCPMDVEGSRVWFMDLWNYSLVPY
ILEAVREGLQMYGKRTPWEDPSKWVLDTYPWSSATLPQESPALLQLRPEDVGYESCTST
KEATTSKHIPQTDTEGDPLMNMLMKLQEAANYSSTQSCDSESTSHHEDILDSSLESTL*

>AB023155 cDNA nucleotide sequence

ctatcactaa actgtcattg aattgtactg cattagaaag gaactcaaat atgtgtgacg 60
gcaatggaca tcttgtcacc tttagttggc ctttttcaat gagttaagca ttatatgtgt 120
gttaccaaaa aattattttt tatagttcag agaaccattt ttgttggatg tgtaatttgg 180
aagttttgtt tacattatgt ccttaggggt tttctttgtt ttaacagcat gcagcttgac 240
agaaatacac tacccaaaaa gggactaaga tatacccccat catctcggca ggccaaccaa 300
gaagagggca aagagtggtt gcgttctcat tctactggag ggcttcagga cactggcaac 360
cagtcacctc tggtttcccc ttctgccatg tcatcttctg cagctggaaa ataccacttt 420
tctaacttgg tgagcccaac aaatttgtct caatttaacc ttcccgggcc cagcatgatg 480
cgctcaaaca gcatcccagc ccaagactct tccttcgatc tctatgatga ctcccagctt 540
tgtgggagtg ccacttctct ggaggaaaga cctcgtgcca tcagtcattc gggctcattc 600
agagacagca tggaagaagt tcatggctct tcattatcac tggtgtccag cacttcttct 660
ctttactcta cagctgaaga aaaggctcat tcagagcaaa tccataaact gcggagagag 720
ctggttgcat cacaagaaaa agttgctacc ctcacatctc agctttcagc aaatgctcac 780
cttgtagcag cttttgaaaa gagcttaggg aatatgactg gccgattgca aagtctaact 840
atgacagcgg aacaaaagga atctgaactt atagaactaa gagaaaccat tgaaatgctg 900
aaggctcaga attctgctgc ccaggcggct attcagggag cactgaatgg tccagaccat 960
cctcccaaag atcttcgcat cagaagacag cattcctctg aaagtgtttc tagtatcaac 1020
agtgccacaa gccattccag tattggcagt ggtaatgatg ccgactccaa gaagaagaaa 1080
aagaaaaact gggtgaactc tagaggaagt gagctgagaa gttctttcaa acaagccttt 1140
gggaagaaaa agtccaccaa gcctccttca tcacattctg acattgaaga gcttactgat 1200
tcatcccttc cggcatcccc caagttaccc cataatgctg gtgactgtgg ctcagcatcc 1260
atgaagccct cacaatctgc ttcagcgatc tgtgaatgca cagaagctga ggcagagata 1320
attctgcagc tgaagagcga gctcagagaa aaggaattaa aattaacgga tattcggctg 1380
gaggccctca gctctgctca tcatcttgat cagatccggg aagccatgaa ccggatgcag 1440
aatgaaattg aaatactgaa agctgaaaat gaccggttga aggcagaaac tggtaacaca 1500
gctaagccta ctcggccacc gtcagaatcc tcaagcagca cctcctcttc atcttccagg 1560
cagtcattag gactttctct aaacaatttg aacatcacag aggctgttag ctcagatatt 1620
ttgctagatg atgctggtga tgcaactgga cataaagatg gccgcagtgt gaaaattata 1680
gtctccataa gcaagggcta tggtcgagca aaggaccaaa aatctcaggc atatttgata 1740
ggatccattg gtgttagtgg aaaaaccaag tgggatgtct tagatggtgt aataagacgt 1800
ctctttaagg aatatgtatt ccgaattgat acatccacta gccttggtct gagctctgac 1860
tgcattgcta gctactgtat aggagactta attagatccc ataacctaga agtgcctgaa 1920
ttgctgcctt gtggatacct tgttggagat aataacatca tcactgtgaa cctcaaaggg 1980
gtagaagaaa atagtttgga cagttttgtt tttgatacgc tgattcctaa accaattacc 2040
caaaggtact ttaacttgtt gatggagcat cacagaatta tactctcagg accgagtggt 2100
actggaaaga cctatttggc aaacaaactt gctgaatatg taataaccaa atctggaagg 2160
aaaaaaacag aggatgcaat tgccactttt aatgtggacc acaagtcaag taaggaattg 2220
caacaatatc tagctaacct ggctgaacag tgcagtgctg ataataatgg agtggagctc 2280
ccagttgtaa taattcttga taatcttcat catgtgggct ctctgagtga tatcttcaat 2340
ggttttctca attgtaaata caacaaatgt ccatatatta ttggaacaat gaatcaggga 2400
gtttcttcat caccaaatct agagctgcat cacaatttca ggtgggtatt atgtgcaaat 2460
catacagaac cagtgaaagg ctttttaggc agatatcttc gaagaaaact catagagata 2520
gaaattgaaa ggaacattcg caataatgac ctagtcaaaa ttatagattg gattccgaag 2580

Figure 1h (CONTINUED)

acgtggcatc atctcaacag ttttttggaa acacacagtt cttctgacgt taccattggt 2640
ccccgactat tccttccttg ccccatggat gtagaaggtt ctagagtatg gttcatggat 2700
ctctggaact attctttagt accttatatt ctggaggcag tgagagaggg tcttcagatg 2760
tatgggaaac gcacaccatg ggaagatcct tcaaagtggg tgcttgacac atatccatgg 2820
agctcagcaa ctctgcctca ggagagccca gccttacttc agctgcgacc agaagatgtt 2880
gggtatgaaa gctgcacatc cactaaggaa gccacaacct caaagcacat tccgcaaact 2940
gacacagaag gagatcccct gatgaatatg ctaatgaaac tccaagaagc agccaattac 3000
tcgagcacac aaagctgcga cagcgaaagc accagccacc atgaagacat tttggattca 3060
tctcttgaat ctaccctcta gagggtgaaa aaagttaagg gaaaagactt tgcttttaaa 3120
aaaatgtttc aaaagaaagg tattttcact aaaccactgc cagtataaaa gcaccctgtc 3180
aagggccctg acccagagtt gtggtctcca aggaggcagc agaactaagt ctgaaccgcc 3240
aagatgctaa attgcaatgg aagcttaact ttagtttatt tctaagcatt ttttatatct 3300
gtggagtaat agaaagctcc attactcaac tggaaaggac cctaatgaca gggcaactga 3360
acagattgca catgggatag ccaaactgga ctttctttgt ttcctcttta aaagtttaca 3420
atgcagacca ttttttgtcc cttccttttg tttcctctga ggggctgttc gccccaggca 3480
gggtccatct ttctgatctg tccaacctcc tttgtgccac acggtgctgg tcacagggct 3540
tcagtagtgt ttgtgttgtg cgctcacccc attccagaac aaatccaaga ggccagtcct 3600
ccataagcac aaatggaatt gtgcaaccac cagaaaaaca ctactgtggc aaactggaga 3660
agtgccaatt taattctaac tgccacgttc tcatgatgtg ctccaccaac tttttagtat 3720
atgagtcact ggttttataa ggttgttttt accacagtgg tctttttaaa ccacctgccc 3780
actcccttaa caagagtttt ataccaatta ttagtcaaca ctgataaaag gctttttag 3840
ggctttattt gtttgagcct tttcagtgaa agaaggaaca tttcctatgg tgctgtctca 3900
ctgccttaaa acagatttct atgacagttt aacagttggt ttaaatccta aaccattggt 3960
aatttccact gtcttttcat ttacaaccaa gcaacaccag ttaacatagt agcctcatct 4020
ctatatatct ttctcttttt ttttttttt tgaagaaatg gataggagaa agatcagtat 4080
ttttagcctt gtgaatagat cgctttgcct atcctccaaa atattaaaat aacccagaaa 4140
tgctctttga ccgtcactta aaacctaaga catgtggcga aattccatcc agttctaagt 4200
gaaagagttt cagaaggcag gagattttga attattatcc agcagggctg gaagcactag 4260
atgcagcatg agcacaacta ttcggctttc cttccctatt gtttttgttt ttttaatgag 4320
ttttgacgca tgttgttttg attgctattg ttgtacatga gaaattcagc attaaagaac 4380
actgaagcgg taaggtcact gtggaagagg aagcgtttat actgtaaaag aaggttagat 4440
ttgcacagtc tactgggtag gtattgtaaa taataatttt taaaacttgc acaaatcaaa 4500
acaaacacaa acaaaattgt attttatcct attggtgtta agaggtgttt cacttgctga 4560
gatttcctgt acattgcaaa caaatacaga atgcaaaccc tcaaagctgt attatctggt 4620
gtgtttgtcc tgtatttaca gttgtttttg actatgcagg agctatcagt gctagagtga 4680
gcatgcttca aaactgtaca tgaagccaat atatttttgg ataagtaaaa ctgtctgaaa 4740
gtacatctgt catggcaggc tttaaagaga gtgcatgaaa actgatcagt cattggagaa 4800
gttaccacca cacacaaagg acaggtttta agtttatgaa acccaagggc taggccatgg 4860
tatagacttc ttctatgagt gtgtgaaaat gtgttacttt taggacgtgt atttggtgct 4920
actctctgtg accaccaatg ggtcagttgc tatagaacaa caacaccacg aaacatctgt 4980
gcagttttca gagtgtcaca aagtcaatag gtccttacac ggtgctattg ccctaaggga 5040
aatccgaact gaatttatgc acatagaatt gtcaccctga ctttgaagcc tcaaacatgg 5100
atcaaatctg ttgtgaaaca tcaatatatg tagctggatg agtgactagt ttcccttgta 5160
taatatgtga tctaagaaaa ttgctaatct ttccctgcca ttttgagaaa cacagtccaa 5220
acatgagcat aaacagaatt tcctgcaata catcccagta ggtccaccta gtttacaact 5280
taaactagtt tgtgaaacat ttgtctgtat acattttata ttttgtacat tttgatgtaa 5340
catatcatgt aaataggcag aaacagtgaa ataaatcatc tgaaaagttt tgtagtcttt 5400
gtaaagcccc aacaataagt acttggtgtc aatggactta actggatgat gtattttcta 5460
ttggtttatt gttcctctag cttgtaaacc agcttgcata tatttttttg caaatgtgca 5520
ccctgtatct gtctaaatta ttactttgcc attaaagtgg aattatttat tgac 5574

Figure 1i. Overview of cloned nematode and human unc-53s variants

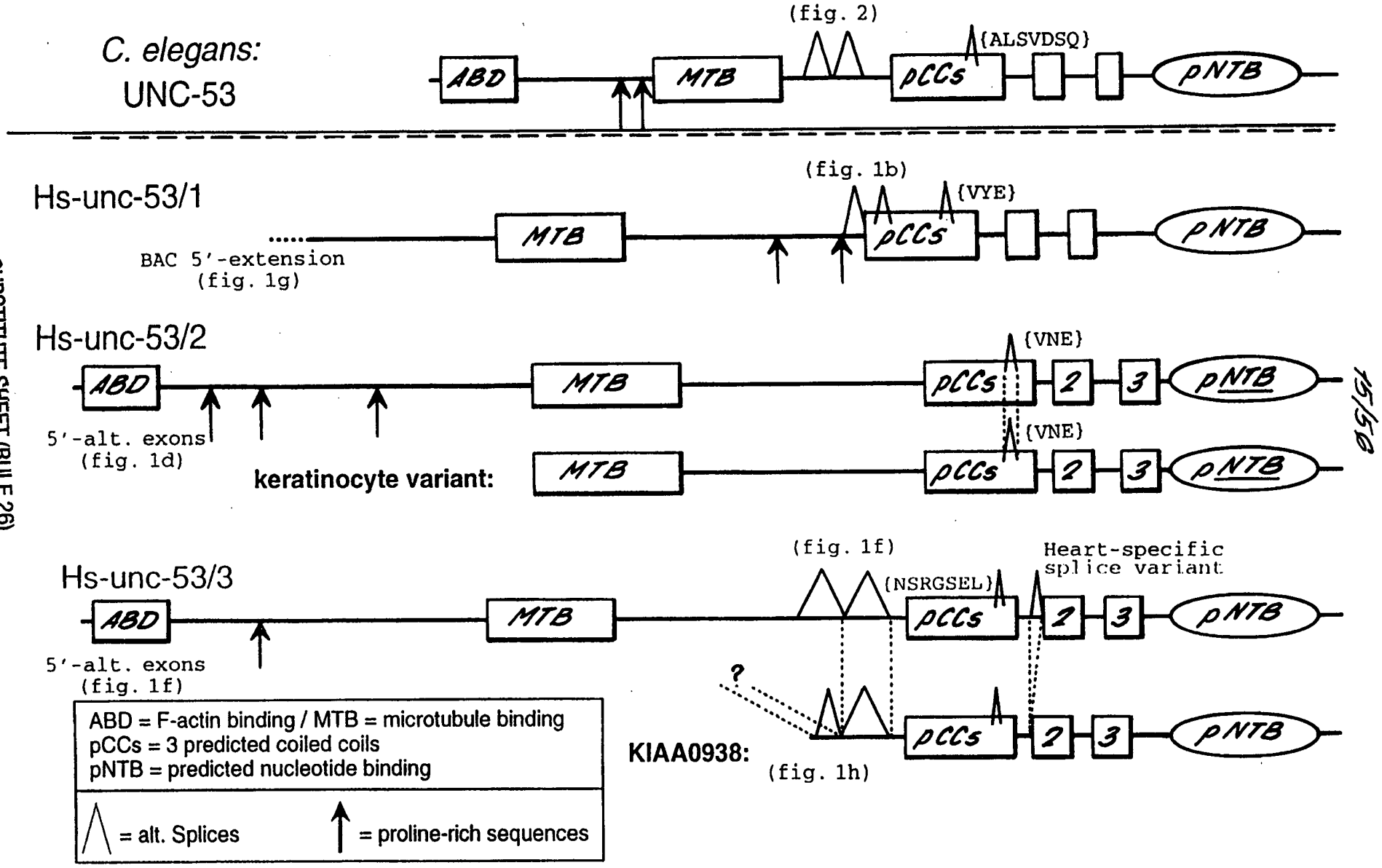


Figure 2: Illustration of a multiple sequence alignment between the different members of the Unc-53 protein family.

```
Ce-unc-53
Hs-unc-53/3     1 MPVLGVASKLRQPAVGSKPVHTALPIPNLGTTGSQHCSSRPLELAETESSMLSCQLALKS
Hs-unc-53/2                                                             MES
Hs-unc-53/1

Ce-unc-53                    MTTSNVELIP.IYTDWANRHLSKGSLSKSIRDISNDFRDYRLVSQLINV
Hs-unc-53/3    61 TCEFGEKKPLQGKAKEKEDSK.IYTDWANHYLARSGHKRLIKDLQQDIADGVLLAEIIQI
Hs-unc-53/2     4 VSESSQQQKRKPVIHGLEDQKRIYTDWANHYLTKSGHKRLIKDLQQDVTDGVLLAQIIQV
Hs-unc-53/1

Ce-unc-53      49 IVPINEFSPAFTKRLAKITSNLDGLETCLDYLKNLGLDCSKLTKTDIDSGNLGAVL
Hs-unc-53/3   120 IA..NEKVEDINGCPRSQSQMIENVDVCLSFLAARGVNVQGLSAEEIRNGNLKAIL
Hs-unc-53/2    64 VA..NEKIEDINGCPKNRSQMIENIDACLNFLAAKGINIQGLSAEEIRNGNLKAIL
Hs-unc-53/1

Ce-unc-53     105 QLLFLLSTYKQKLRQLKKDQKKLEQLPTSIMPPAVSKLPSPRVATSATASAT.........
Hs-unc-53/3   174 GLFFSLSRYKQ...QQHHQQQ..YYQSLVELQQRVTH.ASPPSEASQAKTQQDMQS(SLAA
Hs-unc-53/2   118 GLFFSLSRYKQ...QQQQPQK..QHLS.SPLPPAVSQVAGAPSQCQAGTPQQQVPV.TPQA
Hs-unc-53/1

Ce-unc-53     157 ..NPNSNFP.QMSTSRLQTPQ SRISKIDS..SKIGIKPKTSGLKPPSSSTTSSNNT.NSF
Hs-unc-53/3   226 ..RYATQSNHSGIATSQKKPT)RLPGP..SRVPAAGSSSKVQGA.....SNL..NRRSQSF
Hs-unc-53/2   172 PCQPHQPAPHQQSKAQAEMQS RLSGP.TARVSAAGSEAKTRGG.....STTANNRRSQSF
Hs-unc-53/1

Ce-unc-53     211 R.............PSSR......SSGNNNVGSTISTSA.KSLESSSTYSSISNLNR..
Hs-unc-53/3   277 NSIDKNK....PPNYANGNEKDS.SKGPQSSSG..VNGNVQPPSTAGQ........PPAS
Hs-unc-53/2   226 NNYDKSKPVTSPPPPPSSHEKEPLASSASSHPG..MSDNAPASLESGSS.STPTNCSTSS
Hs-unc-53/1

Ce-unc-53     248 ..PT..SQLQKPSRPQTQLVRVATTTKIGSSK......LAAPKAVSTPKLASVKTI.GAK
Hs-unc-53/3   322 AIPSP.SAS.KPWRSKSMNVKHSATSTMLTVKQSSTATSPTPSS...DRLKP.PVSEGVK
Hs-unc-53/2   283 AIPQPGAAT.KPWRSKSLSVKHSATVSMLSVK.......PPGPEA...PR....PTPEAMK
Hs-unc-53/1

Ce-unc-53     297 QEPDNSGGGGGGML.KLKLFSSKNPSSSSNSP..QPT..RKAAAVP.....QQ.QTLSKI
Hs-unc-53/3   376 TAPSGQK....SMLEKPKLVNARTALRPPQPPSSGPSDGGKDD..DAFSESGEMEGPNSG
Hs-unc-53/2   329 PAPNNQK....SMLEKLKLFNSKGGSKAGEGPGSRDTSCERLETLPSFEESEELEAASRM
Hs-unc-53/1

Ce-unc-53     346 AAPVKSGLKPPTS...KL...GSA.TSMSKLCTPKVS.............YRKTD......
Hs-unc-53/3   430 LNSG..G...STNSSPKVSPKLAPPKAGSKNLSNKKSLLQPKEKE......EKNRDKNK.
Hs-unc-53/2   385 LTTV..G...PASSSPKIALKGIAQRTFSRALTNKKSSLKGNEKEKEKQQREKDKEKSKD
Hs-unc-53/1

Ce-unc-53     381 ..........................................................
Hs-unc-53/3   478 .....VCTEK.PVKEEKDQ.........VTEMAPKKTSKIASLIPKGSKTTAAKKESLIP
Hs-unc-53/2   440 LAKRASVTERLDLKEEPKEDPSG...AAVPEM.PKKSSKIASFIPKGGKLNSAKKEPMAP
Hs-unc-53/1     1          ..MLPKRAKAPGGGGGMAKASAAELKVFKSGSVDSRVPGGPPASNLRKQKSLT

Ce-unc-53     381 ...........APIISQQDSKRCSKSSEEESGYAGFNSTSPTSSSTEGSLM.HSTSSKSS
Hs-unc-53/3   523 SSSGIPKPGSKVPTVKQTISPGSTASKESEKFRTTKGSPSQSLSKP.IT.MEKASASSCP
Hs-unc-53/2   496 SHSGIPKPGMKSMPGKSPSAP..APSKEGERSRSGKLSSGLPQQKPQLDG.RHSSSSSSL
```

Figure 2 (CONTINUED 1)

```
Ce-unc-53     429 TSDE.KSPSSDDLTLNASIVT.AIRQPIAATPVS.PNIIN.......KPVEE..KP.TLA
Hs-unc-53/3   581 APLEGREAGQ..ASPSGS.CTMTVAQSSG..QSTGNG..AVQLP...Q.QQQHSHPNTAT
Hs-unc-53/2   553 ASSEGKGPGG..TTLNHSISSQTVSGSVGTTQTTGSNTVSVQLP...QPQQQYNHPNTAT
Hs-unc-53/1   106 FQAKGSPAGG..A....KTPLAPLAPNLGKPSRIPRGPYAEVKPLSKAPEAAVSEDGKSD

Ce-unc-53     476 VKGVKSTAKKDPPPAV..PPRDTQPTIG..V.VSPIMAHKKLTNDPVISEK.....PEPE
Hs-unc-53/3   630 VAPFIYRAHSENEGTALPSADSCT.S.P..TKMDL..S.YSKTAKQCLEEISGE...GPE
Hs-unc-53/2   608 VAPFLYRSQTDTEGNV..TAESSS.T G..VSVEP..SHFIKTGQPALEELTGE...DPE
Hs-unc-53/1   160 DELLSSKAKAQKSSGPVPSAKGQE.E.RAFLKVDP...ELVVTVLGDLEQLLFSQMLDPE

Ce-unc-53     526 ..KLQSMSIDTTDV.PPLP.PLKSVVPLKMTSIRQP.PTYD.....VLLKQGKI
Hs-unc-53/3   680 TRRMRTVK.NIADLRQNLEETMSSLRGTQISHSTLE.TTFDSTVTTEVNGRTIPNLTSRP
Hs-unc-53/2   657 ARRLRTVK.NIADLRQNLEETMSSLRGTQVTHSTLE.TTFDTNVTTEMSGRSILSLTGRP
Hs-unc-53/1   215 SQRKRTVQ.NVLDLRQNLEETMSSLRGSQVTHSSLEMTCYDSD...DANPRSVSSLSNRS

Ce-unc-53     570 T.....SPVKSFGY
Hs-unc-53/3   738 TPMTWRLGQACPRLQAGDAPSLGAGY.PRSGTSRFIHTDPSRFMYTTPLRRAAVSRLGNM
Hs-unc-53/2   715 TPLSWRLGQSSPRLQAGDAPSMGNGYPPRANASRFINTESGRYVYSAPLRRQLASRGSSV
Hs-unc-53/1   271 SPLSWRYGQSSPRLQAGDAPSVGGSCRSEGTPAWYMHGERAHYSHTMPMR..SPSKLSHI

Ce-unc-53     579 EQSSASEDSIVAHASAQVTPPTKTSGNHSLERRMGKN.KTSE..SSGYTSDAGVAMCAKM
Hs-unc-53/3   797 SQIDMSEKA.SS.DLDMSS.EVDVGGYMSDGDILGKSLRTDD.INSGYMTDGGLNLYTRS
Hs-unc-53/2   775 CHVDVSDKA.GD.EMDLEGISMDAPGYMSDGDVLSKNIRTDD.ITSGYMTDGGLGLYTRR
Hs-unc-53/1   329 SRLELVESLDSD.EVDLKS.....GYMSDSDLMGKTMTEDDDITTG.....

Ce-unc-53     636 REKLKEYDDM..TRRA..QN.....GYPDNFEDSSSLSSGISDNNELDDISTDDLSGV..
Hs-unc-53/3   853 LNRIP..D.TATSRDIIQRGVHDVTVDADSWDDSSSVSSGLSDT..LDNISTDDLNTTSS
Hs-unc-53/2   832 LNRLP..DGMAVVRETLQRNTSLGLGDADSWDDSSSVSSGISDT..IDNLSTDDINTSSS
Hs-unc-53/1   369 .....WDESSSISSGLSDA..SDNLSSEEFNASSS

Ce-unc-53     685 .....D..MATVASKHS.....
Hs-unc-53/3   908 VSSYSNITVPSRQNTQ..LRTDSEKRSTTDET..WDSPEELKKPEE.DFDS...HGDAG.
Hs-unc-53/2   888 ISSYANTPASSRKNLD..VQTDAEKHSQVERNSLW.SGDDVKKSDG.GSDSG.IKMEPG.
Hs-unc-53/1   397 LNSLPSTPTASRRNSTIVLRTDSEKRSLAESGLSWPSESEEKAPKKLEYDSGSLKMEPGT

Ce-unc-53     695 .....
Hs-unc-53/3   959 GKWKTVSSGLPEDPEKA.GQKASLSVSQTGSWRRGMSAQGG..AP.SRQKAGTSALKTP.
Hs-unc-53/2   942 SKWRRNPSDVSDDSDKSTSGKKNPVISQTGSWRRGMTAQVGITMPRTKASAPAGALKTPG
Hs-unc-53/2                     E
Hs-unc-53/1   457 SKWRRERPESCDDSSKGGELKKPISLGHPGSLKKGKTPPVAVTSPITH..TAQSALKV..

Ce-unc-53     695 .....
Hs-unc-53/3  1014 .GKTDDAKASEKGKAPLKGSSLQRSPSDAGKSSGDEGKKP.PSGIGRSTA.TSSFGFKKP
Hs-unc-53/2  1002 TGKTDDAKVSEKGRLSPKASQVKRSPSDAGRSSGDESKKPLPSSSRTPTANANSFGFKKQ
Hs-unc-53/1   513 AGK.PEGKATDKGKLAVKNTGLQRSSSDAGRDRLSDAKKP.PSGIARP.STSGSFGYKKP

Ce-unc-53     695 .....
Hs-unc-53/3  1071 SGVGSS.AMITSSGATITSGSATLGKIPKSAAIGGKSNAGRKTSLDGSQNQDDVVLHVSS
Hs-unc-53/2  1062 SGSAAGLAMITASGVTVTSRSATLGKIPKSSALVSRS.AGRKSSMDGAQNQDDGYLALSS
Hs-unc-53/1   570 P.PATGTATVMQTG.....GSATLSKIQKSSGIPVKPVNGRKTSLDVSNSAEPGFLAPGA

Ce-unc-53     695 .....
Hs-unc-53/3  1130 KTTLQYRSLPRPSKSSTSGIPGR.GGHRSSTSSID.SNVSSKSAGATTSKLREPTKIGSG
Hs-unc-53/2  1121 RTNLQYRSLPRPSKSNSR...NG.AGNRSSTSSID.SNISSKSAGLPVPKLREPSKTALG
Hs-unc-53/1   624 RSNIQYRSLPRPAKSSSMSVTGGRGGPRPVSSSIDPSLLSTKQGGLTPSRLKEPTKVASG
```

Figure 2 (CONTINUED 2)

```
Ce-unc-53     695 ..........................................................
Hs-unc-53/3  1188 .RSS.PVIVNQTDKEKEK......VAVSDSESVSLSGSPKSSPTSASACG.AQGLRQPGS
Hs-unc-53/2  1176 .SSL.PGLVNQTDKEKG........ISSDNESVASCNSVKVNPAAQPVSSPAQTSLQPGA
Hs-unc-53/1   684 .RTT.PAPVNQTDREKEKAKAKAVALDSDNISLKSIGSPESTPKN.......QASHPTAT

Ce-unc-53     695 ..... ..............................DYSHFVRHPTSSSSKPRVP
Hs-unc-53/3  1239 KYPDIASPTFRR(LFGAKAGGKSASAPNTEGVKSSSVMPSPSTTLARQGSLESPSSGTGSM
Hs-unc-53/2  1226 KYPDVASPTLRR LFGGKP.TKQVPIATAENMKNSVVISNPHATMTQQGNLDSP.SGSGVL
Hs-unc-53/1   735 KLAELPPTPLRA T..AKSFVKPPSLANLDKVNSN.............SLDLPSS.....

Ce-unc-53     714 SRSSTSVDSRSRAEQENVYKLLSQCRTSQRGAAATSTFGQHSLRSPG........YSSYS
Hs-unc-53/3  1299 GSAGGLSGSSSPLFNKPSDLTTDVISLSHSLAS....SPASVHSFTSGGLVWAANMSSSS
Hs-unc-53/2  1284 S ....SGSSSPLYSKNVDLN......QSPLAS....SPSSAHSAPSNSLTWGTNASSSS
Hs-unc-53/1   774 ........SDTTHASKVPDLHATSSASGGPLPSCFTPSPAPILNINSASFSQGLELMSGP

Ce-unc-53     766 PHLSVSADKDTMS.MHSQTSRRPSSQKPSYSG(CFHSLDRKCHLQEF.TSTEHRMAALLSP
Hs-unc-53/3  1355 AGSKDTPSYQSMTSLHTSSESIDLPLS..... HHGSLSGLTTGTHEVQSL..LMR.TGSV
Hs-unc-53/2  1329 AVSKDGLGFQSVSSLHTSCESIDISLSSGGVP SHNSSTGLIASSKD.DSLTPFVR.TNSV
Hs-unc-53/1   826 SVPKETRMYPKLSGLHRSMESLQMPMS..... ........LPSAFPSSTPVPTPP.APPA

Ce-unc-53     824 RRVPNSMS KYDSS)(AAALNASGMSRSMILLESL SPRPPRRHQSPADS CIITASPSAPRRS
Hs-unc-53/3  1407 RSTL.SES).....  ...................(MQLDRNTLPKKGLR)YTPSSRQANQEEG
Hs-unc-53/2  1387 KTTL.SES .PLSS  PAASPKFCRSTLPRKQDSD PHLDRNTLPKKGLR YTPTSQLRTQEDA
Hs-unc-53/1   872 APTE.EET .EELT  WSGSPR.......AGQLDS NQRDRNTLPKKGLR Y....QLQSQEET

Ce-unc-53     883 HSPRGPTARIPLSL..ASSPVHVNNNW)GSYSARSRGGSST GIYGETF.............
Hs-unc-53/3  1441 KEWLRSHSTGGLQDTGNQSPLVSPSAM SSSAAGKYHFSNL VSPTNLSQFNLPGPSMMRSN
Hs-unc-53/2  1444 KEWLRSHSAGGLQDTAANSPFSSGSSV TSPSGTRFNFSQL ASPTTVTQMSLSNPTMLRTH
Hs-unc-53/1   918 KERRHSHTIGGLPESDDQSELPSPPAL PMSLSAKGQLTNI(VSPTAAT.....TPRITRSN

Ce-unc-53     928 ..............QLHRLS...DEKSPAHSAKSEM.......  ..GSQLSLASTT..AY
Hs-unc-53/3  1501 SIPAQDSSFDLYDDSQLCGSATSLEERPRAIS.HSGSFRDSMEE VHGSSLSLVSSTSSLY
Hs-unc-53/2  1504 SLSNADGQYDPYTDSRFRNSSMSLDEKSRTMS.RSGSFRDGFEE VHGSSLSLVSSTSSVY
Hs-unc-53/1   973 SIPTHEAAFELYSGSQM.GSTLSLAERPKGMI.RSGSFRDPTDD)VHGSVLSLASSASSTY

Ce-unc-53     959 GS LNEKYEEA .IRDMARDLECYKNTVDSLTKKQ.... .....................
Hs-unc-53/3  1560 ST AEEKAHSE QIHKLRRELVASQEKVATLT.SQLSAN .....................
Hs-unc-53/2  1563 ST PEEKCQSE .IRKLRRELDASQEKVSALT.TQLTAN .....................
Hs-unc-53/1  1031 SS(AEERMQSE)QIRKLRRELESSQEKVATLT.SQLSAN(VSAMKYGKIKAVLITIVRQVQPR

Ce-unc-53     991 .ENYG A.LFDLFEQKLRKLTQHIDRSNLKPEEAIRFRQDIAHLRDISNHLASNSAHANEGAG
Hs-unc-53/3  1596 ..... AHLVAAFERSLGNMTGRLQSLTMTAEQK...ESELIELRETIEMLKAQNSAAQAAIQ
Hs-unc-53/2  1598 ..... AHLVAAFEQSLGNMTIRLQSLTMTAEQK...DSELNELRKTIELLKKQNAAAQAAIN
Hs-unc-53/1  1090 EENYL)ANLVAAFEQSLVNMTSRLRHLAETAEEK...DTELLDLRETIDFLKKKNSEAQAVIQ

Ce-unc-53    1041 ELL.....................RQPSLESVASHRSSMSSSSKSSKQEKISLSSFGKNR
Hs-unc-53/3  1650 GALNGPDHPPR........DLRIRRQHSSESVSSINSATSHSSIGS...GNDADSKKKKK
Hs-unc-53/2  1652 GVINTPELNCKGNGTAQSADLRIRRQHSSDSVSSINSATSHSSVGS...NIESDSKKKKR
Hs-unc-53/1  1149 GALNASETTPK........ELRIRRQNSSDSISSLNSITSHSSIGS...SKDADAKKKKK

Ce-unc-53    1090 KSW(ALSVDSQ)IRSSLSK.FTKKKN.K....NYD.......EAHMPS...I..S.GSQG..
Hs-unc-53/3  1699 KNW(VNSRGSE)LRSSFKQAFGKKKSTKPPSSHSDIEELT..DSSLPASPKLPHNAGDCGSA
Hs-unc-53/2  1709 KNW(VN....E)LRSSFKQAFGKKKSPKSASSHSDIEEMT..DSSLPSSPKLPHN.GSTGS.
Hs-unc-53/1  1198 KSW(V....YE)LRSSFNKAFSIKKGPKSASSYSDIEEIATPDSSAPSSPKLQHGSTETASP

Ce-unc-53    1129 ..... T.L.....DN..... ....ID.....VIELKQELKERDSALY
```

Figure 2 (CONTINUED 3)

```
Ce-unc-53    1151 EVRLDNLDRAREVDVLRETVNKLKTENKQLKKEVDKL...TNGP..ATRASSRAS...I.
Hs-unc-53/3  1816 DIRLEALSSAHHLDQIREAMNRMQNEIEILKAENDRLKAETGNTAKPTRPPSESSSSTSS
Hs-unc-53/2  1800 DIRLEALSSAHQLDQLREAMNRMQSEIEKLKAENDRLKSESQGSG.CSRAPSQVS...IS
Hs-unc-53/1  1310 DIRLEALNSAHQLDQLRETMHNMQLEVDLLKAENDRLKVAPGPSSGST..PGQVPGSSAL

Ce-unc-53    1202 ..PVIYD...DEHVYDAACSSTS. ......ASQSSKRSSGCNSIKVTVNV..DIAGEI SS
Hs-unc-53/3  1876 SSSR.QSLGLSLNNLN.ITEAVSS DILLDDAGDATGHKD.GRSVKIIVSISKGYGRAK DQ
Hs-unc-53/2  1856 ASPR.QSMGLSQHSLN.LTESTSL(DMLLDDTGECSARKEGGRHVKIVVSFQEEMKWKE)DS
Hs-unc-53/1  1368 SSPR.RSLGLALTHSF.GPSLADT DLSPMDGISTCGPKEE.VTLRVVVRMPPQHIIKG DL

Ce-unc-53    1248 IVNPDKEIIVGYLAMS    TSQSCWKDI.DVSILGLFEVYLSRIDVEHQLGIDARDSILGYQI
Hs-unc-53/3  1933 ..KSQA.YLIGSIGVS    .GKTKW.DVLDGVIRRLFKEYVFRIDTSTSLGLSS.DCIASYCI
Hs-unc-53/2  1914 ..RPHL.FLIGCIGVS(*).GKTKW.DVLDGVVRRLFKEYIIHVDPVSQLGLNS.DSVLGYSI
Hs-unc-53/1  1425 ..KQQE.FFLGCSKVS    .GKVDW.KMLDEAVFQVFKDYISKMDPASTLGLST.ESIHGYSI

Ce-unc-53    1307 GELRRVIGDSTTMITSH..PTDILT.SSTTIRMFMHGAAQSRVDSLVLDMLLPKQMILQL
Hs-unc-53/3  1987 GDLIR....SHNLEVPELLPCGYLVGDNNIITVNLKGVEENSLDSFVFDTLIPKPITQRY
Hs-unc-53/2  1968 GEIKR....SNTSETPELLPCGYLVGENTTISVTVKGLAENSLDSLVFESLIPKPILQRY
Hs-unc-53/1  1479 SHVKR....VLDAEPPEMPPCRR..GVNN.ISVSLKGLKEKCVDSLVFETLIPKPMMQHY

Ce-unc-53    1364 VKSILTERRLVLAGATGIGKSKLAKTLAAYVSIRTNQS.EDSIV.NISIPENNKEELLQ
Hs-unc-53/3  2042 FNLLMEHHRIILSGPSGTGKTYLANKLAEYVITKSGRKKTEDAIATFNVDHKSSKELQQ
Hs-unc-53/2  2023 VSLLIEHRRIILSGPSGTGKTYLANRLSEYIVLREGRELTDGVIATFNVDHKSSKELRQ
Hs-unc-53/1  1531 ISLLLKHRRLVLSGPSGTGKTYLTNRLAEYLVERSGREVTEGIVSTFNMHQQSCKDLQL

Ce-unc-53    1421 VERRLEKILRSKESC.....IVILDNIPKNRIAFVVSVFANVPLQN...NEGPFVVCTVN
Hs-unc-53/3  2102 YLANLAEQCSADNNGVELPVVIILDNL..HHVGSLSDIF.NGFL.NCKYNKCPYIIGTMN
Hs-unc-53/2  2083 YLSNLADQCNSENNAVDMPLVIILDNL..HHVSSLGEIF.NGLL.NCKYHKCPYIIGTMN
Hs-unc-53/1  1591 YLSNLANQIDRETGIGDVPLVILLDDL..SEAGSISELV.NGAL.TCKYHKCPYIIGTTN

Ce-unc-53    1473 R..YQIPELQIHHNFKMSVMSNRLE...GFILRYLRRRAVEDEYRLTVQMPSELFKII
Hs-unc-53/3  2158 QGVSSSPNLELHHNFRWVLCANHTEPVKGFLGRYLRRKLIEIEIERNIRNN.DLVKII
Hs-unc-53/2  2139 QATSSTPNLQLHHNFRWVLCANHTEPVKGFLGRFLRRKLMETEISGRVRNM.ELVKII
Hs-unc-53/1  1647 QPVKMTPNHGLHLSFRMLTFSNNVEPANGFLVRYLRRKLVESDSDINANKE.ELLRVL

Ce-unc-53    1526 DFFPIALQAVNNFIEKTNSVDVTVGPRACLNCPLTVDGSREWFIRLWNENFIPYLERVA
Hs-unc-53/3  2215 DWIPKTWHHLNSFLETHSSSDVTIGPRLFLPCPMDVEGSRVWFMDLWNYSLVPYILEAV
Hs-unc-53/2  2196 DWIPKVWHHLNRFLEAHSSSDVTIGPRLFLSCPIDVDGSRVWFTDLWNYSIIPYLLEAV
Hs-unc-53/1  1704 DWVPKLWYHLHTFLEKHSTSDFLIGPCFFLSCPIGIEDFRTWFIDLWNNSIIPYLQEGA

Ce-unc-53    1585 RDGKKTFGRCTSFEDFTDIVSKKWPWFDGENPEN....VLKRLQLQDL......VPSPAN
Hs-unc-53/3  2274 REGLQMYGKRTPWEDPSKWVLDTYPW..SSATLPQESFALLQLRPEDVGYESCTSTKEAT
Hs-unc-53/2  2255 REGLQLYGRRAPWEDPAKWVMDTYPW..AASPQQHEWPPLLQLRPEDVGFDGYSMPREGS
Hs-unc-53/1  1763 KDGIKVHGQKAAWEDPVEWVRDTLPW..PSA..QQDQSKLYHLFPPTVGPHSIASPPEDR

Ce-unc-53    1635 SSRQ......HFNPL.ESLIQL.HATKH...QTIDNI
Hs-unc-53/3  2332 TSKHIPQTDTEGDPLMNMLMKLQEAANYSSTQSCDSE..STSHHEDILDSSLESTL
Hs-unc-53/2  2313 TSKQNPPSDAEGDPLMNMLMRLQEAANYSSPQSYDSDSNSNSHHDDILDSSLESTL
Hs-unc-53/1  1819 TVKDSTPSSLDSDPLMAMLLKLQEAANY..IESPDRET.........ILDPNLQATL
```

Figure 3: Illustration of a multiple sequence alignment between C. elegans Unc-53 (Ce) and C. Briggsae Unc-53 (Cb).

```
Cb   1 MTTSNVELIPIYTDWANRHLSKGALSRPIRDISNEFRDYRLVSQLINVIVPINEYSPTYTYPLAKITSNLDGLETCLDYL
Ce   1 MTTSNVELIPIYTDWANRHLSKGSLSKSIRDISNDFRDYRLVSQLINVIVPINEFSPAFTKRLAKITSNLDGLETCLDYL

Cb  81 KNLGLDCSKLTKTDIDSGNLGAVLQLLFLLSSYKQKLRQLKKDQKKLEQLFVTITTTTAIMPPAVSNIPYSRLPSPRVPPA
Ce  61 KNLGLDCSKLTKTDIDSGNLGAVLQLLFLLSTYKQKLRQLKKDQKKLEQLF.....TSIMPPAVSKLFSPRVATSATASA

Cb161 SNPNSNFTQMSTSRLQTPQSRISKPDSTKIGIKPKTTSGLRPP.STTSSNTNINSFRPSSRSSGNNNVGSTISTSARSLD
Ce156 TNPNSNFPQMSTSRLQTPQSRISKIDSSKIGIKPK.TSGLKPPSSSTTSSNNTNSFRPSSRSSGNNNVGSTISTSAKSLE

Cb260 SSSAYSSISNLSRPTPSSQIQKPTSRLQTQQVRVATTTKIGSSKLAAPKAVSTPKLASVKTIKTTTTEHDNS....SGML
Ce235 SSSTYSSISNLNRPT..SQLQKP.SRPQTQLVRVATTTKIGSSKLAAPKAVSTPKLASVKTI.GAKQEPDNSGGGGCGML

Cb316 KLKLFSSKNASSSNNSPQPLRKA ....EQ..SKIAAPVKIGLKPPTSSTNKLGSATSMSKLCTPKVSYRKPDTLLHTKS
Ce311 KLKLFSSKNPSSSSNSPQPTRKAAAVPQQQTLSKIAAPVKSGLKPPTS...KLGSATSMSKLCTPKVSYRKTDAPIISQQ

Cb389 DSKRCSKSSEEESGYAGFNSTSPASSSTEGSLSMHSTSSKSSTSDEKSPSSDDLTLNASIVTAIRQPIATSAVSP.VISK
Ce388 DSKRCSKSSEEESGYAGFNSTSPTSSSTEGSLSMHSTSSKSSTSDEKSPSSDDLTLNASIVTAIRQPIAATPVSPNIINK

Cb468 PVEEKPTLAVKGV.SASPKLPPPTVTERTNQPTIGVVSPIMAHKKLTNDPVISEK..PEPEKLQSMSIDCTDVPPLP.PL
Ce468 PVEEKPTLAVKGVKSTAKKDPPPAVPPRDTQPTIGVVSPIMAHKKLTNDPVISEK..PEPEKLQSMSIDCTDVPPLP.PL

Cb546 KSLERVPPKMTPIRQPPTYDVLVKQGKITSPVKSFGYDQVESSASEDSIVAH..VQMAPPVQKTSAGQSSMERRIQKNKT
Ce545 KSV..VPLKMTSIRQPPTYDVLLKQGKITSPVKSFGYEQ..SSASEDSIVAHASAQVTPPT.KTS.GNHSLERRMGKNKT

Cb624 SESSGYASDAGVAMCAKMREKLKEYTDMTRRAQNGYPDNPEDSSSLSSGISDNNELDDISTDDLSGIDMATVASKHSDYS
Ce619 SESSGYTSDAGVAMCAKMREKLKEYTDMTRRAQNGYPDNPEDSSSLSSGISDNNELDDISTDDLSGVDMATVASKHSDYS

Cb704 HFVRHTSSSSSRPRVPSRPSISVDSRSRVEQENVYKLLSQCRTSQRGAATATSSFGQHSLRSPGYSSYSPHLTVSADKDT
Ce699 HFVRHPTSSSSKPRVPSRSSTSVDSRGKAEQENVYKLLSQCRTSQRGAA.ATSTFGQHSLRSPGYSSYSPHLSVSADKDT

Cb784 MSMHSQTSRRPSSQKPSYAGQFHSLDRKCHLQEFTSAEHRMAALLSPRRVPNSMSKYDSSSGSYSARSRGGSSTGIYGEP
Ce778 MSMHSQTSRRPSSQKPSYSGQFHSLDRKCHLQEFTSTEHRMAALLSPRRVPNSMSKYD.ESGSYSARSRGGSSTGIYGET

Cb864 FQLHRLSDEKSPAHSARSEMGSQLSLASTTAYGTLNEKYEHAIRDMARDLECYKNTVDSLTKKQENYGALFDLFEQKLRK
Ce857 FQLHRLSDEKSPAHSAKSEMGSQLSLASTTAYGSLNEKYEHAIRDMARDLECYKNTVDSLTKKQENYGALFDLFEQKLRK

Cb944 LTSHIDRSNLKPDEATRFRQDIAHLREISNHLATNSMHVNEGAGELLRQPSLESVASHRSSMSSSSKSSKQEKISLSSFG
Ce937 LTQHIDRSNLKPEEATRFRQDIAHLRDISNHLASNSAHANEGAGELLRQPSLESVASHRSSMSSSSKSSKQEKISLSSFG

Cb1024 KNKKSWIRSSLSKFTKKKNKNYDEGHMPSISGSQGTLDNIDVIELRQELKERDSALYEVRLDNLDRAREVDVLKETVNKL
Ce1017 KNKKSWIRSSLSKFTKKKNKNYDEAHMPSISGSQGTLDNIDVIELKQELKERDSALYEVRLDNLDRAREVDVLRETVNKL

Cb1104 KNENKQLKKEVDKLTNTSITRASSRASLPYIQDDEHVYDHACSSTSASQSSKRSSGCNSIKVTVNVDIAGEISSIVNPDR
Ce1097 KTENKQLKKEVDKLTNGPATRASSRASIPVIYDDEHVYDAACSSTSASQSSKRSSGCNSIKVTVNVDIAGEISSIVNPDK

Cb1184 EIIVGYLPMPANNSTWKDICDSILDSPEKYLSKIDLDRQLGLDAKDAIFGYQIGELRRVIGDSSTIITSHPVDILTPTTT
Ce1177 EIIVGYLAMSTSQSCWKDIDVSILGLFEVYLSRIDVEHQLGIDARDSILGYQIGELRRVIGDSTTMITSHPTDILTSSTT

Cb1264 IRMFMYGAAQSRVDSMVLDMLLPRQMILQLVKSIVTERRLVLAGATGIGKSKLAKTLAAYVSLQTNQSEDKIVNITIPEN
Ce1257 IRMFMHGAAQSRVDSLVLDMLLPKQMILQLVKSILTERRLVLAGATGIGKSKLAKTLAAYVSIRTNQSEDSIVNISIPEN

Cb1344 NKEELLQVERRLEKILRSKEACVVILDNIPKNRIAFVVSVFANVPLQNNEGPFVVCTVNRYQIPELKTIPNFKMSVMSNR
Ce1337 NKEELLQVERRLERILRSKESCIVILDNIPKNRIAFVVSVFANVPLQNNEGPFVVCTVNRYQIPELQIHHNFKMSVMSNR

Cb1424 LEGFILRYLRRRAVEDEYRLSVQMPSELSRIIEFFPVALQAVNNFIEKTNSVDVTVGPRACLNCPLTIDGSREWFIRLWN
Ce1417 LEGFILRYLRRRAVEDEYRLTVQMPSELFKIIDFFPIALQAVNNFIEKTNSVDVTVGPRACLNCPLTVDGSREWFIRLWN

Cb1504 QNFIPYMERVARDGKKTLGRCTSFEDPTDIVTEKWPWFDGPNPEDVLKRLQLQDLAPSPANSSRQPFNPLESLIQLHATK
Ce1497 ENFIPYLERVARDGKKTFGRCTSFEDPTDIVSKKWPWFDGENPEDVLKRLQLQDLVFSPANSSRQHFNPLESLIQLHATK
```

Figure 4. Prosite Signatures

Block A, Large family:

IYTDWANXXLX(K,R)(A,G,S,T)XXX(K,R)X(ILVA)(H,K,R,T,S)D(I,L)XXDXXDXXL(L,V
)(A,S)(N,D,Q,E)(I,L,V,A)I(N,D,Q,E)(I,L,V,A)(I,L,V,A)(V,A,T,S)X(17,19)(
I,L,F)(N,D,Q,E)X(I,L,V,A)(N,D,Q,E)XCLXXLXXX(A,G,S,T)(I,L,V,A)X(4,5)(I,
L,V,A)(S,T)XX(N,D,Q,E)IXXGXLXA(V,I)LXL(L,F)FXLSX(Y,F)KQ

Block B, Vertebrate:

PEXXRXRTV(Q,K)N(I,L,V,A)(I,L,V,A)DLRQNLEETMSSLRG(S,T)Q(V,I)(S,T)HS(S,T
)LEX(0,1)T

Block C, Vertebrate:

RX(S,T)P(L,M)(S,T)WRXGQ(S,A)XPRLQAGDAPS

Block D, Vertebrate:

GYMSDXD(M,L,V,I)(M,L,V,I)(A,G,S,T)KXXXD(2,3)I(N,T)(A,G,S,T)G(Y,-)

Block E, Vertebrate:

WD(D,E)SSS(M,L,V,I)SSG(L,I)SDXXDN(L,I)S(S,T)(D,E)(D,E)XN(A,G,S,T)(S,T)
SS

Block F, Vertebrate:

DRNTLPKKGLRY

Block G, Large family:

GSX(I,L,V,A)SL(I,L,V,A)S(A,G,S,T)(A,G,S,T)S(0,2)XY(A,G,S,T)XX(E,N)E(K,
R)X(4,5)I(R,H)X(L,M)XR(D,E)LXXXXXXVXXLTXXXXXXXXLXXXFE(Q,K)(S,K)LXXXTXX
(L,I)XX(L,S)XXXXE(Q,E)X(3,6)(D,E)(L,I)XXLRXXX(N,D,Q,E)XLXXXX(A,S)XA(N,
D,Q,E)XXXXXX(L,I)X(0,21)RQXSX(N,D,Q,E)S(I,V)XSXXSXXSXSSX(A,G,S,T)S

Block G, Vertebrate:

SGSFRDXX(D,E)(E,D)VHGSXLSL(V,A)SS(T,A)SSXYS(T,S)XEE(K,R)XXSE(Q,-
)I(R,H)KLRRELX(A,S)SQEKVX(T,A)LT(T,S)QL(S,T)ANAXLVAAFE(Q,K)SLXN(M,L,V,
I)MTXRL(Q,R)XLXXTAE(Q,E)RXXELXXLRXTI(D,E)XLKXXN(A,S)XAQAXIXGX(L,I)N(A,
G,S,T)X(N,D,Q,E)XXXKX(0,8)(N,D,Q,E)LRI(K,R)RQXSS(N,D,Q,E)S(I,V)SS(I,L)
NSXTSHSSXGS

Block H, Large family:

(V,L)DSXVX(D,E)XL(I,L)PKX(M,L,V,I)XXXXXXX(L,I)(M,L,V,I)XXXR(I,L)(I,V)L
(A,S)GX(T,S)GXGK(T,S)XL(A,T)XXLXXY(M,L,V,I)XX(R,K)

and

P(E,N)XX(I,L)HXXF(K,R)XXX(A,S)NXXEX(0,3)GF(L,I)XR(Y,F)L(K,R)(K,R)(K,R)
X(M,L,V,I)(D,E)

and

F(I,L)EXXX(T,S)X(D,E)XXXGPXXX(L,I)XCP(M,L,V,I)X(V,I)(D,E)XX(R,K)XWFXXL
WNXXX(I,V)PY(L,I)XXX(A,V)(R,K)(D,E)GXXXXGXX(T,A)X(F,Y,W)EDP

Block H, Vertebrate:

(V,L)DSXVF(D,E)(T,S)LIPKP(M,L,V,I)XQXYXXLL(M,L,V,I)XHXR(I,L)(I,V)LSGPS
GTGKTYL(A,T)NRLXEY(M,L,V,I)XX(R,K)GR

and

VI(I,L)LD(D,N)LXXXXS(I,L)XX(I,L)XNGXLXCKYXKCPYIIGT(T,M)NQXXXX(T,S)PNXX
LHXXFRXXXX(A,S)NXXEP(A,V)XGFLXR(Y,F)L(K,R)(K,R)(K,R)L(M,L,V,I)(D,E)

and

(R,K)(V,I)(L,I)DWXPKXWXH(I,L)XXFLEXHS(T,S)SDXXIGPXXFLXCP(M,L,V,I)X(V,I

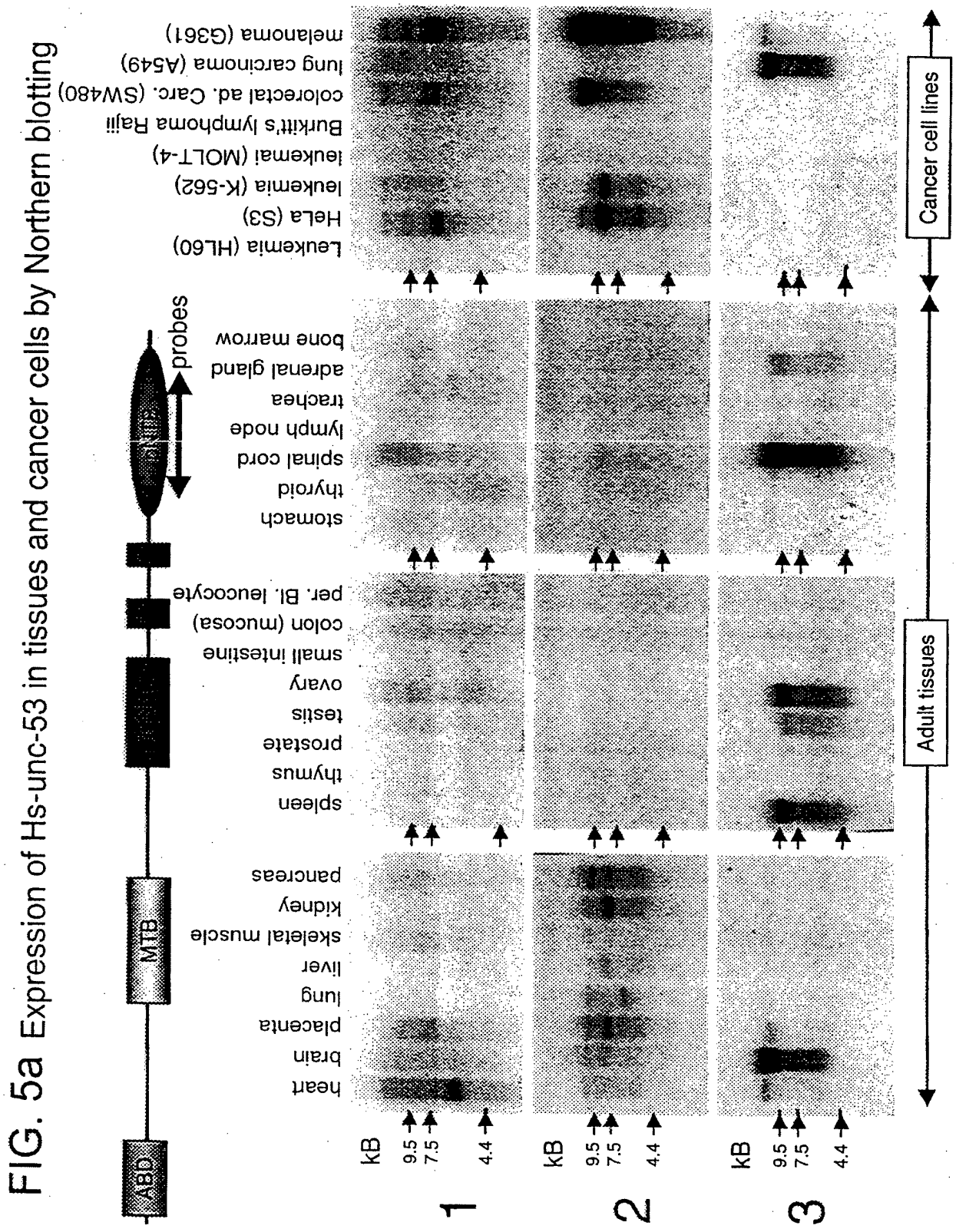


FIG. 5a Expression of Hs-unc-53 in tissues and cancer cells by Northern blotting

FIG. 5b Differential expression of *Hs-unc-53/3* in human brain regions

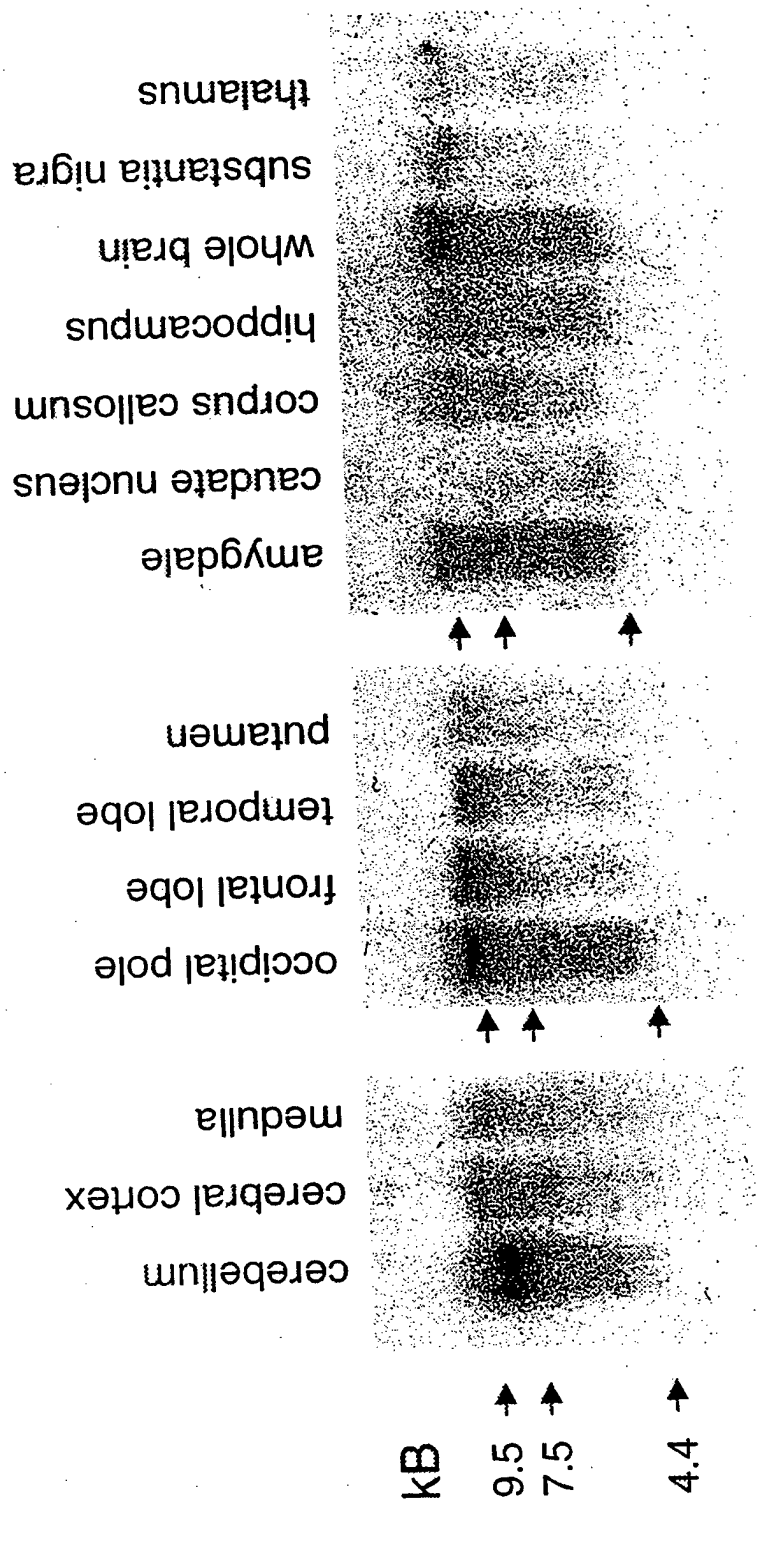
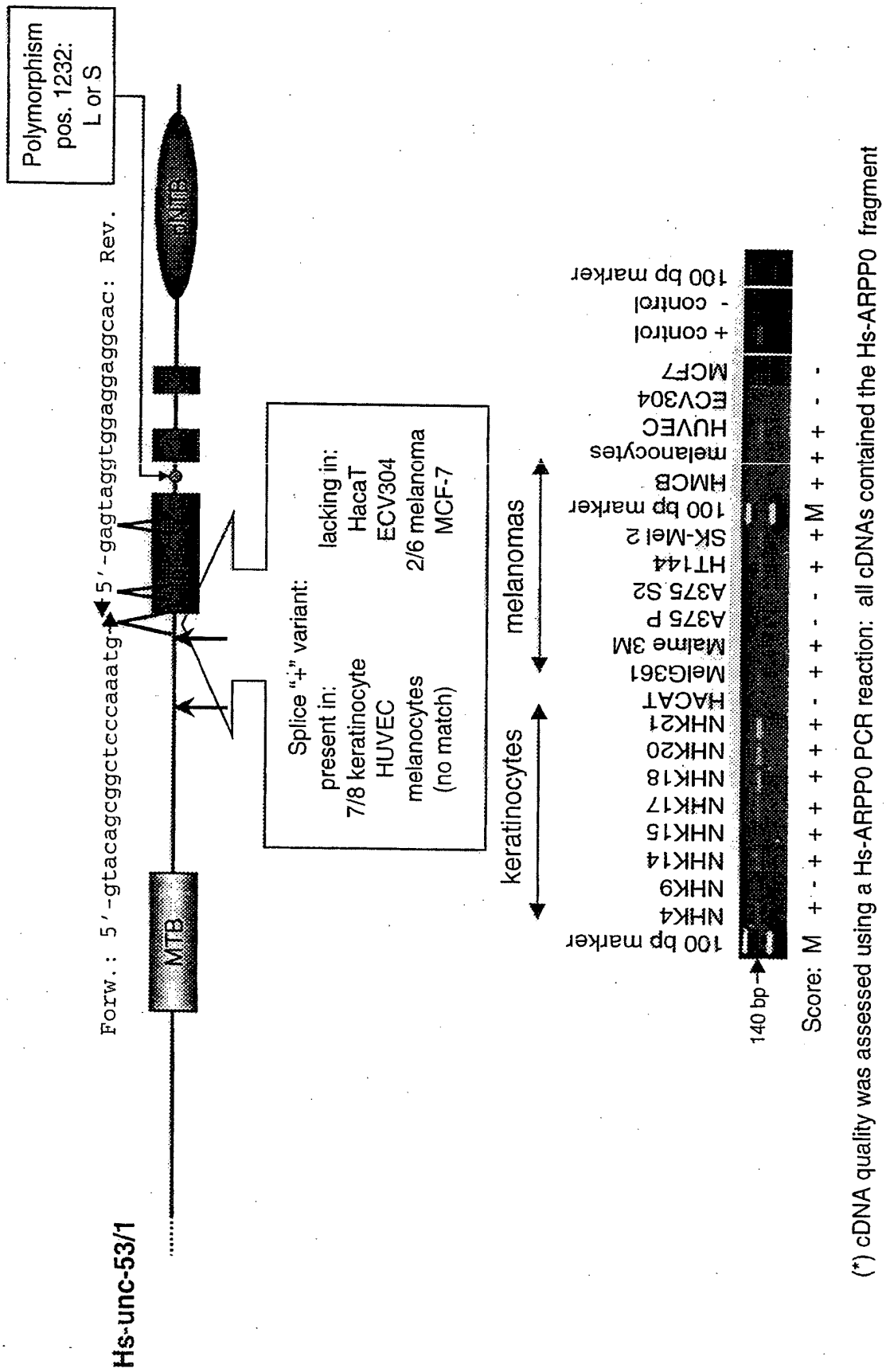
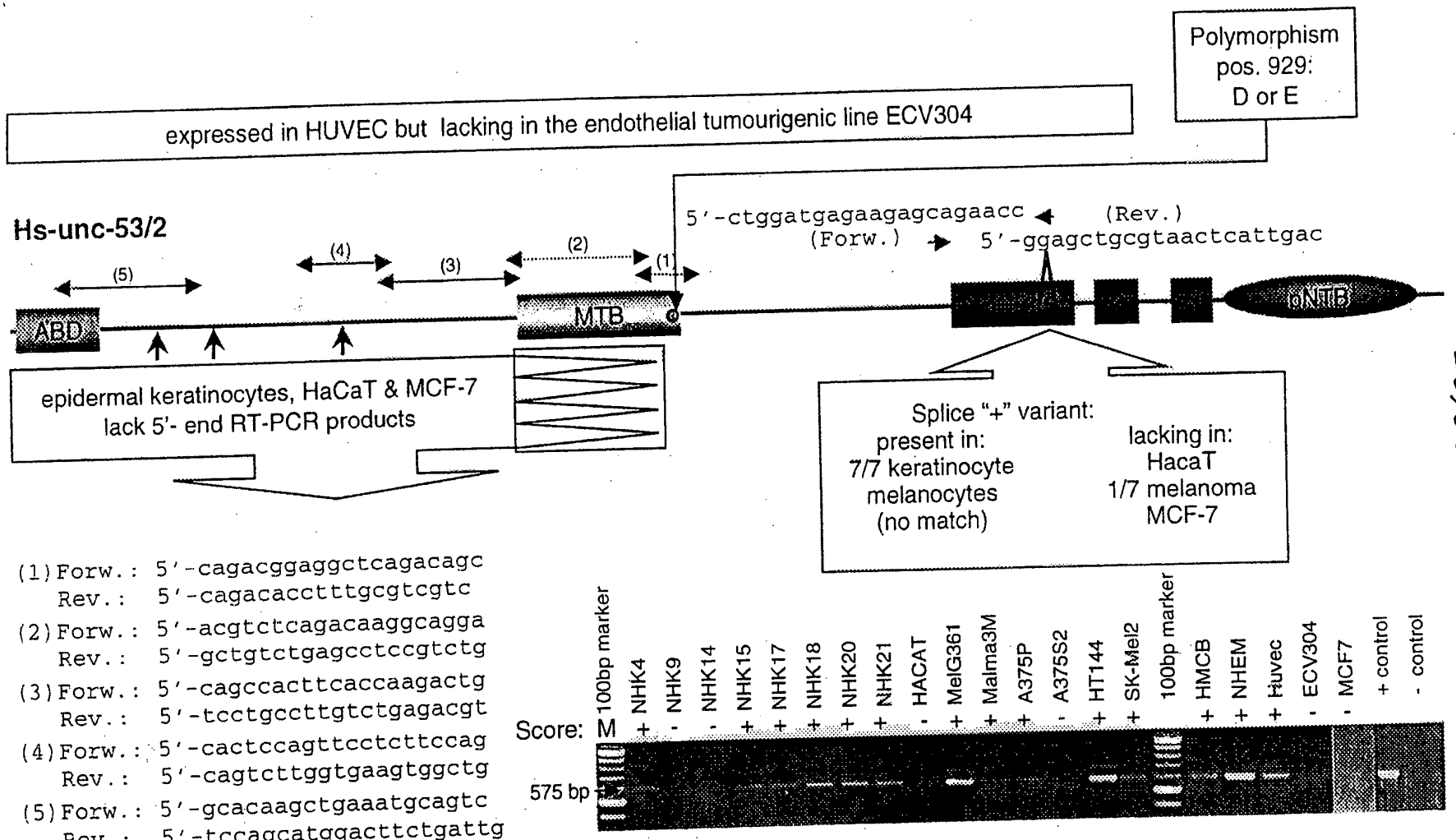


FIG. 6a Hs-unc-53/1 expression: RT-PCR studies*



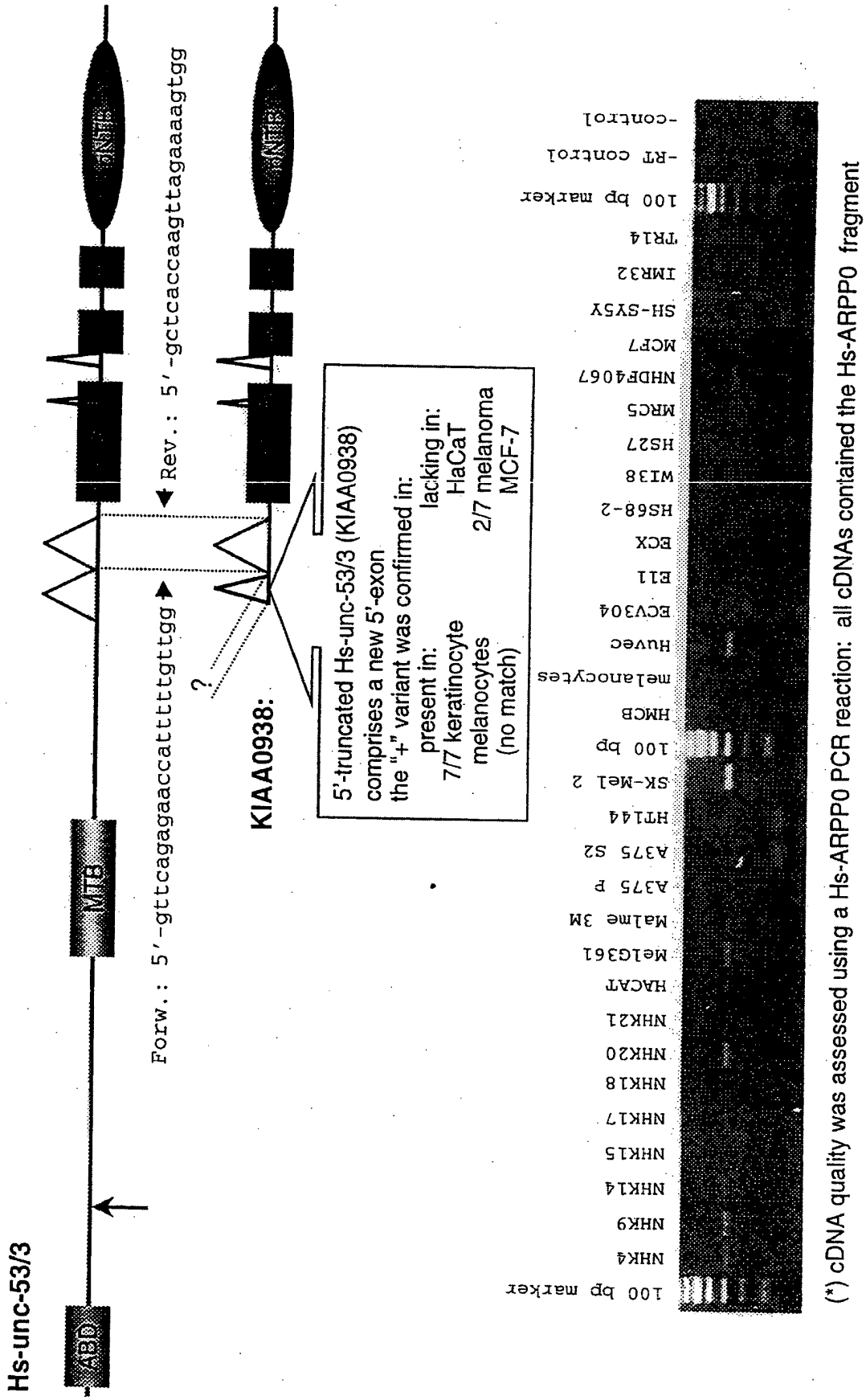
(*) cDNA quality was assessed using a Hs-ARPP0 PCR reaction: all cDNAs contained the Hs-ARPP0 fragment

FIG. 6b Hs-unc-53/2 expression: RT-PCR studies*



(*) cDNA quality was assessed using a Hs-ARPP0 PCR reaction: all cDNAs contained the 650 bp Hs-ARPP0 fragment

FIG. 6c Expression of AB023155 (KIAA0938)



(*) cDNA quality was assessed using a Hs-ARPP0 PCR reaction: all cDNAs contained the Hs-ARPP0 fragment

```
                                     10         20         30         40         50         60         70         80         90         100        110        120
16  1...120                      ********** ********** ********** ****------ ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
 1  1...120 UNC53_KIAA_PCR01 .GTTCAGAGAA CCATTTTTGT TGGATGTGTA ATTTGGAAGT TTTGTTTACA TTATGTCCTT AGGGGTTTTC TTTGTTTTAA CAGCATGCAG CTTGACAGAA ATACACTACC CAAAAAGGGA
 2  1...120 UNC53_KIAA_PCR02 .GTTCAGAGAA CCATTTTTGT TGGATGTGTA ATTTGGAAGT TTTGTTTACA TTATGTCCTT AGGGGTTTTC TTTGTTTTAA CAGCATGCAG CTTGACAGAA ATACACTACC CAAAAAGGGA
 3  1...120 UNC53_KIAA_PCR03 .GTTCAGAGAA CCATTTTTGT TGGATGTGTA ATTTGGAAGT TTTGTTTACA TTATGTCCTT AGGGGTTTTC TTTGTTTTAA CAGCATGCAG CTTGACAGAA ATACACTACC CAAAAAGGGA
 4  1...120 UNC53_KIAA_PCR04 .GTTCAGAGAA CCATTTTTGT TGGATGTGTA ATTTGGAAGT TTTGTTTACA TTATGTCCTT AGGGGTTTTC TTTGTTTTAA CAGCATGCAG CTTGACAGAA ATACACTACC CAAAAAGGGA
 5  1...120 UNC53_KIAA_PCR05 .GTTCAGAGAA CCATTTTTGT TGGATGTGTA ATTTGGAAGT TTTGTTTACA TTATGTCCTT AGGGGTTTTC TTTGTTTTAA CAGCATGCAG CTTGACAGAA ATACACTACC CAAAAAGGGA
 6  1...120 UNC53_KIAA_PCR06 .                                    GGAAGT TTTGTTTACA TTATGTCCTT AGGGGTTTTC TTTGTTTTAA CAGCATGCAG CTTGACAGAA ATACACTACC CAAAAAGGGA
 7  1...120 UNC53_KIAA_PCR07 .GTTCAGAGAA CCATTTTTGT TGGATGTGTA ATTTGGAAGT TTTGTTTACA TTATGTCCTT AGGGGTTTTC TTTGTTTTAA CAGCATGCAG CTTGACAGAA ATACACTACC CAAAAAGGGA
 8  1...120 UNC53_KIAA_PCR08 .GTTCAGAGAA CCATTTTTGT TGGATGTGTA ATTTGGAAGT TTTGTTTACA TTATGTCCTT AGGGGTTTTC TTTGTTTTAA CAGCATGCAG CTTGACAGAA ATACACTACC CAAAAAGGGA
 9  1...120 UNC53_KIAA_PCR09 .    AGAGAA CCATTTTTGT TGGATGTGTA ATTTGGAAGT TTTGTTTACA TTATGTCCTT AGGGGTTTTC TTTGTTTTAA CAGCATGCAG CTTGACAGAA ATACACTACC CAAAAAGGGA
10  1...120 UNC53_KIAA_PCR10 .GTTCAGAGAA CCATTTTTGT TGGATGTGTA ATTTGGAAGT TTTGTTTACA TTATGTCCTT AGGGGTTTTC TTTGTTTTAA CAGCATGCAG CTTGACAGAA ATACACTACC CAAAAAGGGA
11  1...120 UNC53_KIAA_PCR11 .              TTTTTGT TGGATGTGTA ATTTGGAAGT TTTGTTTACA TTATGTCCTT AGGGGTTTTC TTTGTTTTAA CAGCATGCAG CTTGACAGAA ATACACTACC CAAAAAGGGA
12  1...120 UNC53_KIAA_PCR12 .GTTCAGAGAA CCATTTTTGT TGGATGTGTA ATTTGGAAGT TTTGTTTACA TTATGTCCTT AGGGGTTTTC TTTGTTTTAA CAGCATGCAG CTTGACAGAA ATACACTACC CAAAAAGGGA
13  1...120 UNC53_KIAA_PCR13 .GTTCAGAGAA CCATTTTTGT TGGATGTGTA ATTTGGAAGT TTTGTTTACA TTATGTCCTT AGGGGTTTTC TTTGTTTTAA CAGCATGCAG CTTGACAGAA ATACACTACC CAAAAAGGGA
15  1...120 UNC53_KIAA_PCR14 .GTTCAGAGAA CCATTTTTGT TGGATGTGTA ATTTGGAAGT TTTGTTTACA TTATGTCCTT AGGGGTTTTC TTTGTTTTAA CAGCATGCAG CTTGACAGAA ATACACTACC CAAAAAGGGA

                                     130        140        150        160        170        180        190        200        210        220        230        240
16  121...240                    ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
 1  121...240 UNC53_KIAA_PCR01 .CTAAGATATA CCCCATCATC TCGGCAGGCC AACCAAGAAG AGGGCAAAGA GTGGTTGCGT TCTCATTCTA CTGGAGGGCT TCAGGACACT GGCAACCAGT CACCTCTGGT TTCCCCTTCT
 2  121...240 UNC53_KIAA_PCR02 .CTAAGATATA CCCCATCATC TCGGCAGGCC AACCAAGAAG AGGGCAAAGA GTGGTTGCGT TCTCATTCTA CTGGAGGGCT TCAGGACACT GGCAACCAGT CACCTCTGGT TTCCCCTTCT
 3  121...240 UNC53_KIAA_PCR03 .CTAAGATATA CCCCATCATC TCGGCAGGCC AACCAAGAAG AGGGCAAAGA GTGGTTGCGT TCTCATTCTA CTGGAGGGCT TCAGGACACT GGCAACCAGT CACCTCTGGT TTCCCCTTCT
 4  121...240 UNC53_KIAA_PCR04 .CTAAGATATA CCCCATCATC TCGGCAGGCC AACCAAGAAG AGGGCAAAGA GTGGTTGCGT TCTCATTCTA CTGGAGGGCT TCAGGACACT GGCAACCAGT CACCTCTGGT TTCCCCTTCT
 5  121...240 UNC53_KIAA_PCR05 .CTAAGATATA CCCCATCATC TCGGCAGGCC AACCAAGAAG AGGGCAAAGA GTGGTTGCGT TCTCATTCTA CTGGAGGGCT TCAGGACACT GGCAACCAGT CACCTCTGGT TTCCCCTTCT
 6  121...240 UNC53_KIAA_PCR06 .CTAAGATATA CCCCATCATC TCGGCAGGCC AACCAAGAAG AGGGCAAAGA GTGGTTGCGT TCTCATTCTA CTGGAGGGCT TCAGGACACT GGCAACCAGT CACCTCTGGT TTCCCCTTCT
 7  121...240 UNC53_KIAA_PCR07 .CTAAGATATA CCCCATCATC TCGGCAGGCC AACCAAGAAG AGGGCAAAGA GTGGTTGCGT TCTCATTCTA CTGGAGGGCT TCAGGACACT GGCAACCAGT CACCTCTGGT TTCCCCTTCT
 8  121...240 UNC53_KIAA_PCR08 .CTAAGATATA CCCCATCATC TCGGCAGGCC AACCAAGAAG AGGGCAAAGA GTGGTTGCGT TCTCATTCTA CTGGAGGGCT TCAGGACACT GGCAACCAGT CACCTCTGGT TTCCCCTTCT
 9  121...240 UNC53_KIAA_PCR09 .CTAAGATATA CCCCATCATC TCGGCAGGCC AACCAAGAAG AGGGCAAAGA GTGGTTGCGT TCTCATTCTA CTGGAGGGCT TCAGGACACT GGCAACCAGT CACCTCTGGT TTCCCCTTCT
10  121...240 UNC53_KIAA_PCR10 .CTAAGATATA CCCCATCATC TCGGCAGGCC AACCAAGAAG AGGGCAAAGA GTGGTTGCGT TCTCATTCTA CTGGAGGGCT TCAGGACACT GGCAACCAGT CACCTCTGGT TTCCCCTTCT
11  121...240 UNC53_KIAA_PCR11 .CTAAGATATA CCCCATCATC TCGGCAGGCC AACCAAGAAG AGGGCAAAGA GTGGTTGCGT TCTCATTCTA CTGGAGGGCT TCAGGACACT GGCAACCAGT CACCTCTGGT TTCCCCTTCT
12  121...240 UNC53_KIAA_PCR12 .CTAAGATATA CCCCATCATC TCGGCAGGCC AACCAAGAAG AGGGCAAAGA GTGGTTGCGT TCTCATTCTA CTGGAGGGCT TCAGGACACT GGCAACCAGT CACCTCTGGT TTCCCCTTCT
13  121...240 UNC53_KIAA_PCR13 .CTAAGATATA CCCCATCATC TCGGCAGGCC AACCAAGAAG AGGGCAAAGA GTGGTTGCGT TCTCATTCTA CTGGAGGGCT TCAGGACACT GGCAACCAGT CACCTCTGGT TTCCCCTTCT
15  121...240 UNC53_KIAA_PCR14 .CTAAGATATA CCCCATCATC TCGGCAGGCC AACCAAGAAG AGGGCAAAGA GTGGTTGCGT TCTCATTCTA CTGGAGGGCT TCAGGACACT GGCAACCAGT CACCTCTGGT TTCCCCTTCT

                                     250        260        270        280        290        300        310        320        330        340        350        360
16  241...291                    ---------- ---------- ---------- ---------- ---------- -
 1  241...291 UNC53_KIAA_PCR01 .GCCATGTCAT CTTCTGCAGC TGGAAAATAC CACTTTTCTA ACTTGGTGAG C
 2  241...291 UNC53_KIAA_PCR02 .GCCATGTCAT CTTCTGCAGC TGGAAAATAC CACTTTTCTA ACTTGGTGAG C
 3  241...291 UNC53_KIAA_PCR03 .GCCATGTCAT CTTCTGCAGC TGGAAAATAC CACTTTTCTA ACTTGGTGAG C
 4  241...291 UNC53_KIAA_PCR04 .GCCATGTCAT CTTCTGCAGC TGGAAAATAC CACTTTTCTA ACTTGGTGAG C
 5  241...291 UNC53_KIAA_PCR05 .GCCATGTCAT CTTCTGCAGC TGGAAAATAC CACTTTTCTA ACTTGGTGAG C
 6  241...288 UNC53_KIAA_PCR06 .GCCATGTCAT CTTCTGCAGC TGGAAAATAC CACTTTTCTA ACTTGGTG
 7  241...291 UNC53_KIAA_PCR07 .GCCATGTCAT CTTCTGCAGC TGGAAAATAC CACTTTTCTA ACTTGGTGAG C
 8  241...291 UNC53_KIAA_PCR08 .GCCATGTCAT CTTCTGCAGC TGGAAAATAC CACTTTTCTA ACTTGGTGAG C
 9  241...287 UNC53_KIAA_PCR09 .GCCATGTCAT CTTCTGCAGC TGGAAAATAC CACTTTTCTA ACTTGGT
10  241...291 UNC53_KIAA_PCR10 .GCCATGTCAT CTTCTGCAGC TGGAAAATAC CACTTTTCTA ACTTGGTGAG C
11  241...291 UNC53_KIAA_PCR11 .GCCATGTCAT CTTCTGCAGC TGGAAAATAC CACTTTTCTA ACTTGGTGAG C
12  241...291 UNC53_KIAA_PCR12 .GCCATGTCAT CTTCTGCAGC TGGAAAATAC CACTTTTCTA ACTTGGTGAG C
13  241...291 UNC53_KIAA_PCR13 .GCCATGTCAT CTTCTGCAGC TGGAAAATAC CACTTTTCTA ACTTGGTGAG C
15  241...291 UNC53_KIAA_PCR14 .GCCATGTCAT CTTCTGCAGC TGGAAAATAC CACTTTTCTA ACTTGGTGAG C
```

Figure 6d.

Figure 7a. (1) Strategy for cloning 1-2 kb Hu-unc-53/3 fragments

Schematic:

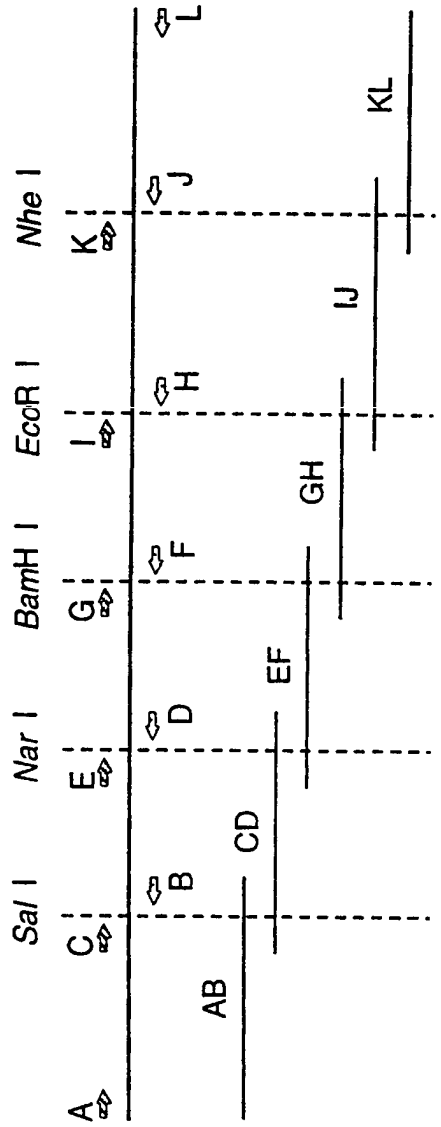


Figure 7a. (2) Strategy for cloning the 3' end of Hs-unc-53/3

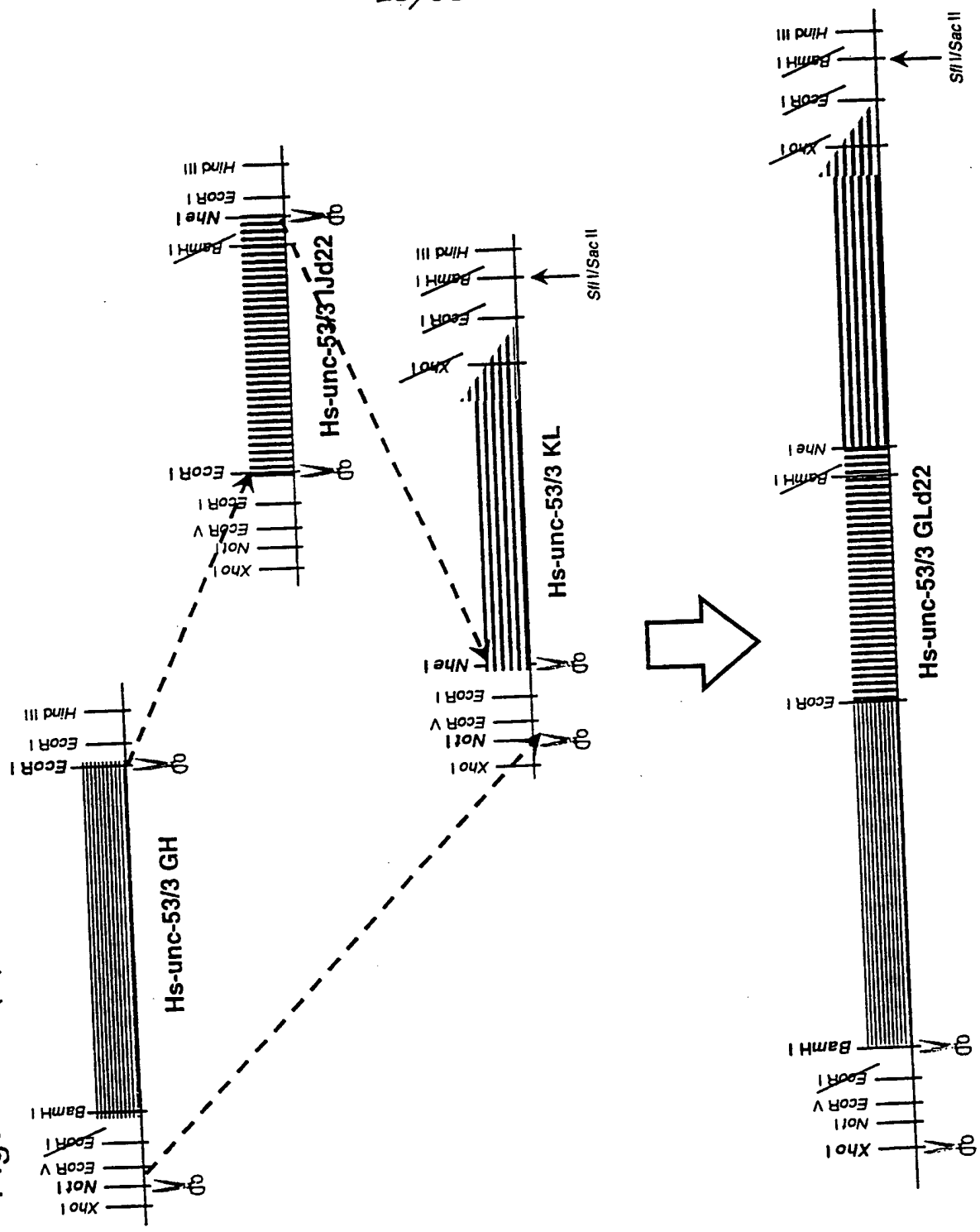


Figure 7a. (3) Strategy for cloning the 5' end of Hs-unc-53/3

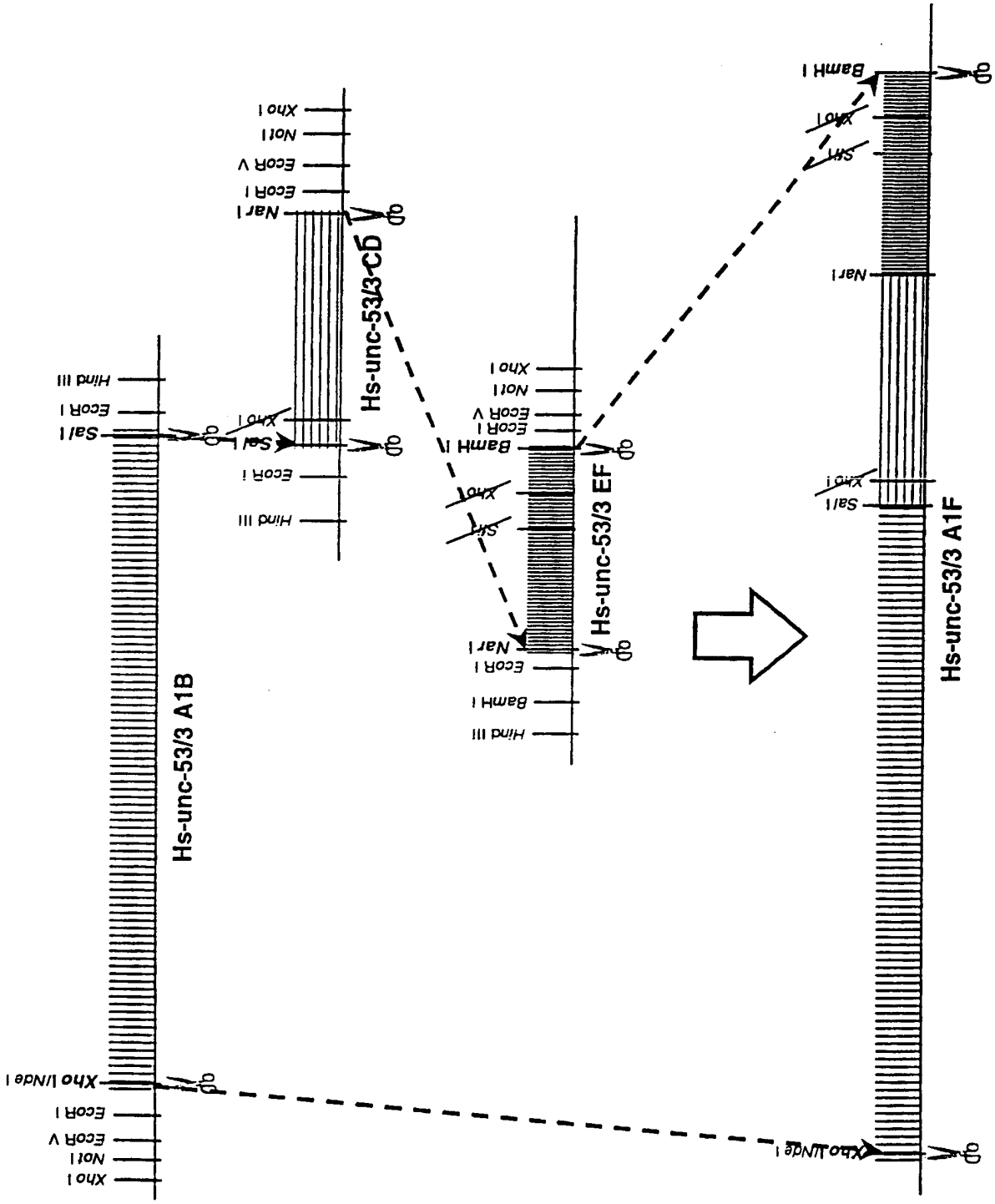
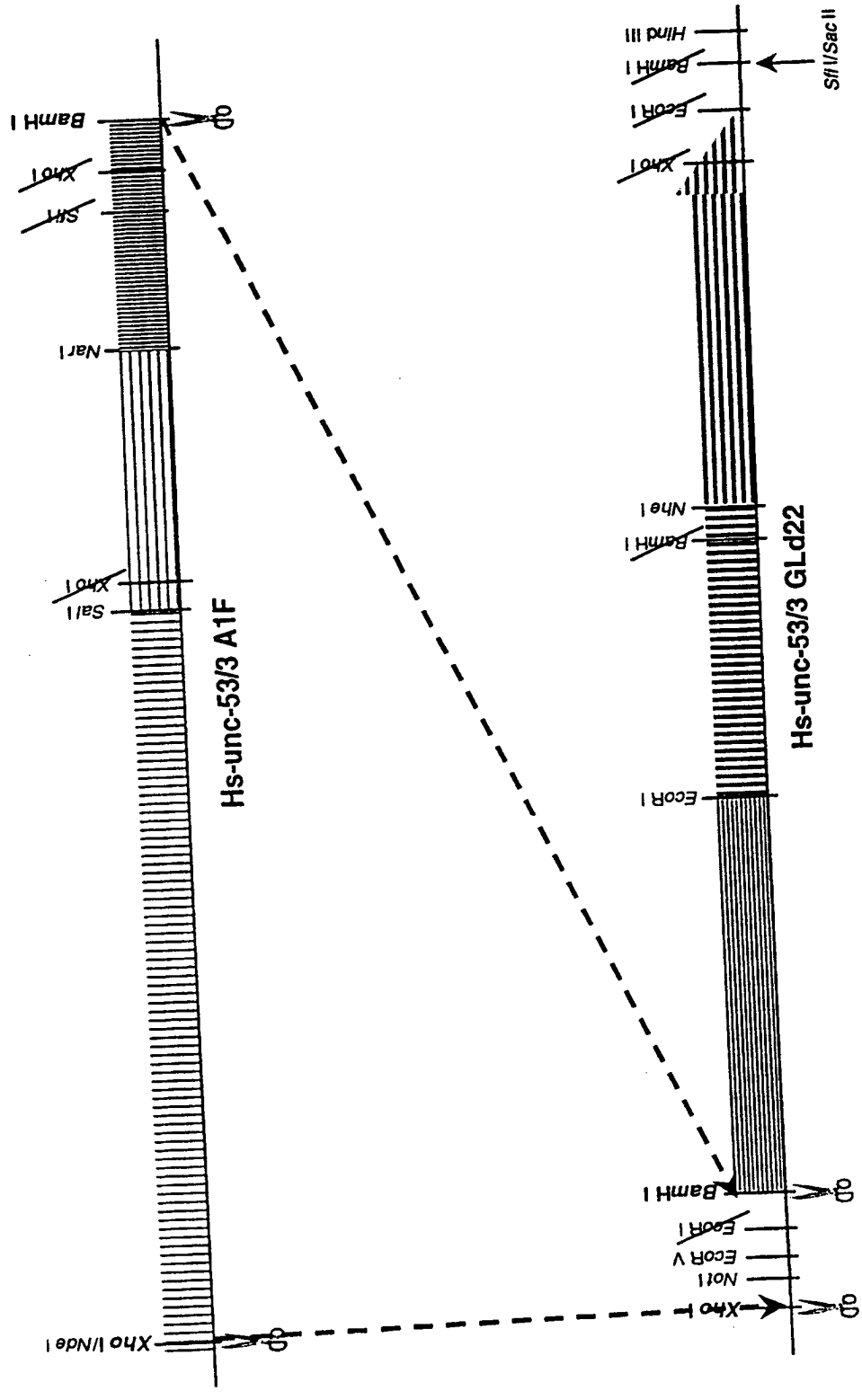


Figure 7a. (4) Strategy for cloning the full-length Hs-unc-53/3 construct



32/56

Figure 7a. (5) Cloning of the Hs-unc-53/3-A1L d22 variant

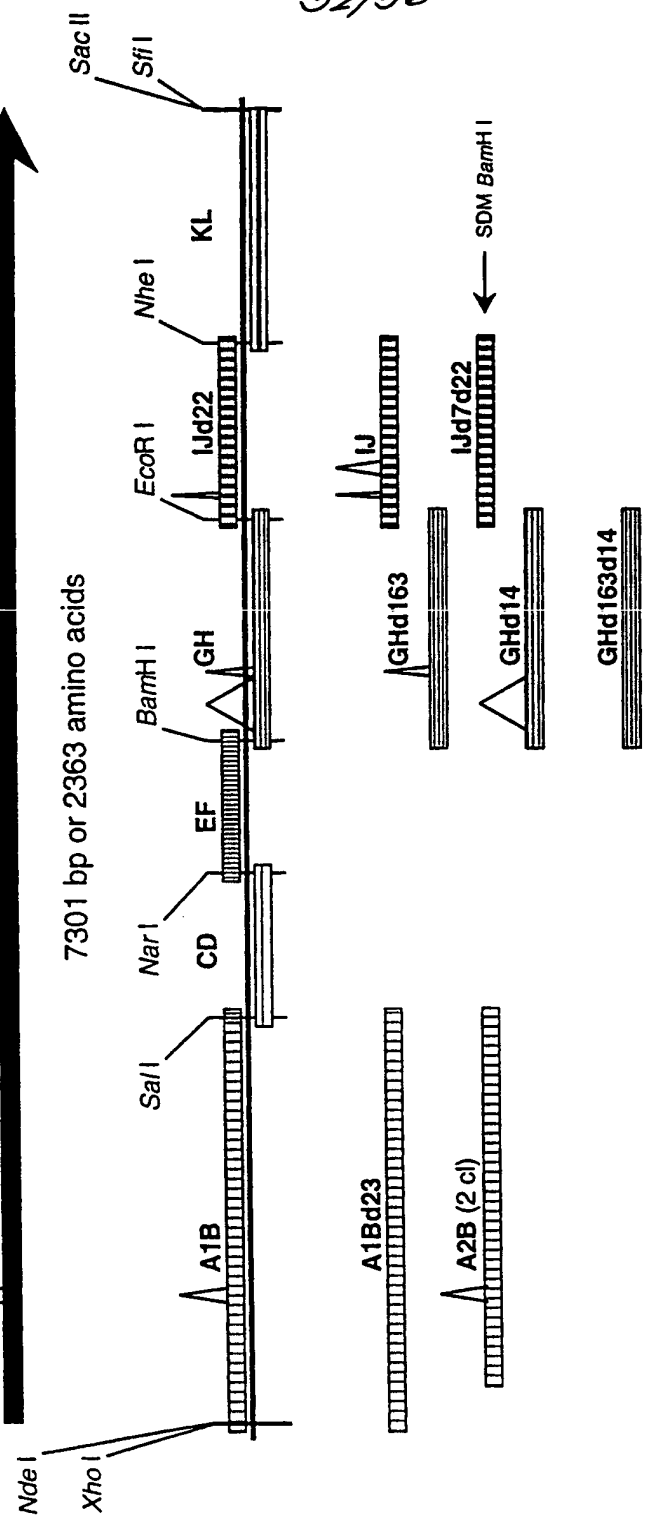
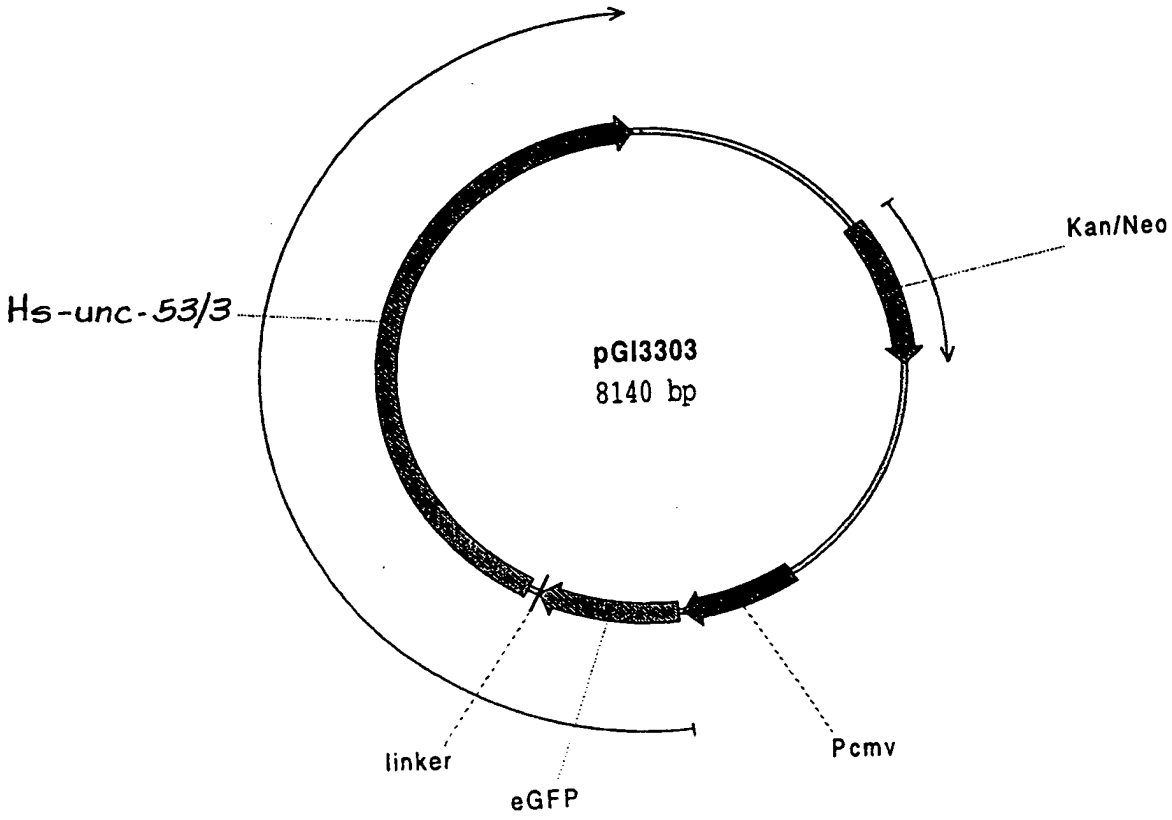


Figure 7b: Illustration of the plasmid map and the nucleotide sequence of the pGI3303 expression vector (C-terminal Hs-unc-53/3 fragment in fusion with GFP)




```
ID   pGI3303                circular DNA; 8140 BP
FT   CDS             1225..2019
FT                   /vntifkey="4"
FT                   /label=Kan/Neo
FT   CDS             3942..4658
FT                   /vntifkey="4"
FT                   /label=eGFP
FT   CDS             4719..8102
FT                   /vntifkey="4"
FT                   /label=Hs-unc-53/3
FT   CDS             4659..4718
FT                   /vntifkey="4"
FT                   /label=linker
FT   promoter        3330..3918
FT                   /vntifkey="29"
FT                   /label=Pcmv
SQ   SEQUENCE   8140 BP;
     CTAGATAACT GATCATAATC AGCCATACCA CATTTGTAGA GGTTTTACTT GCTTTAAAAA      60
     ACCTCCCACA CCTCCCCCTG AACCTGAAAC ATAAAATGAA TGCAATTGTT GTTGTTAACT     120
     TGTTTATTGC AGCTTATAAT GGTTACAAAT AAAGCAATAG CATCACAAAT TTCACAAATA     180
     AAGCATTTTT TTCACTGCAT TCTAGTTGTG GTTTGTCCAA ACTCATCAAT GTATCTTAAC     240
     GCGTAAATTG TAAGCGTTAA TATTTTGTTA AAATTCGCGT TAAATTTTTG TTAAATCAGC     300
     TCATTTTTTA ACCAATAGGC CGAAATCGGC AAAATCCCTT ATAAATCAAA AGAATAGACC     360
     GAGATAGGGT TGAGTGTTGT TCCAGTTTGG AACAAGAGTC CACTATTAAA GAACGTGGAC     420
     TCCAACGTCA AAGGGCGAAA AACCGTCTAT CAGGGCGATG GCCCACTACG TGAACCATCA     480
     CCCTAATCAA GTTTTTTGGG GTCGAGGTGC CGTAAAGCAC TAAATCGGAA CCCTAAAGGG     540
     AGCCCCCGAT TTAGAGCTTG ACGGGGAAAG CCGGCGAACG TGGCGAGAAA GGAAGGGAAG     600
     AAAGCGAAAG GAGCGGGCGC TAGGGCGCTG GCAAGTGTAG CGGTCACGCT GCGCGTAACC     660
     ACCACACCCG CCGCGCTTAA TGCGCCGCTA CAGGGCGCGT CAGGTGGCAC TTTTCGGGGA     720
     AATGTGCGCG GAACCCCTAT TTGTTTATTT TTCTAAATAC ATTCAAATAT GTATCCGCTC     780
     ATGAGACAAT AACCCTGATA AATGCTTCAA TAATATTGAA AAAGGAAGAG TCCTGAGGCG     840
```

Figure 7b (CONTINUED 1)

```
GAAAGAACCA GCTGTGGAAT GTGTGTCAGT TAGGGTGTGG AAAGTCCCCA GGCTCCCCAG      900
CAGGCAGAAG TATGCAAAGC ATGCATCTCA ATTAGTCAGC AACCAGGTGT GGAAAGTCCC      960
CAGGCTCCCC AGCAGGCAGA AGTATGCAAA GCATGCATCT CAATTAGTCA GCAACCATAG     1020
TCCCGCCCCT AACTCCGCCC ATCCCGCCCC TAACTCCGCC CAGTTCCGCC CATTCTCCGC     1080
CCCATGGCTG ACTAATTTTT TTTATTTATG CAGAGGCCGA GGCCGCCTCG GCCTCTGAGC     1140
TATTCCAGAA GTAGTGAGGA GGCTTTTTTG GAGGCCTAGG CTTTTGCAAA GATCGATCAA     1200
GAGACAGGAT GAGGATCGTT TCGCATGATT GAACAAGATG GATTGCACGC AGGTTCTCCG     1260
GCCGCTTGGG TGGAGAGGCT ATTCGGCTAT GACTGGGCAC AACAGACAAT CGGCTGCTCT     1320
GATGCCGCCG TGTTCCGGCT GTCAGCGCAG GGGCGCCCGG TTCTTTTTGT CAAGACCGAC     1380
CTGTCCGGTG CCCTGAATGA ACTGCAAGAC GAGGCAGCGC GGCTATCGTG GCTGGCCACG     1440
ACGGGCGTTC CTTGCGCAGC TGTGCTCGAC GTTGTCACTG AAGCGGGAAG GGACTGGCTG     1500
CTATTGGGCG AAGTGCCGGG GCAGGATCTC CTGTCATCTC ACCTTGCTCC TGCCGAGAAA     1560
GTATCCATCA TGGCTGATGC AATGCGGCGG CTGCATACGC TTGATCCGGC TACCTGCCCA     1620
TTCGACCACC AAGCGAAACA TCGCATCGAG CGAGCACGTA CTCGGATGGA AGCCGGTCTT     1680
GTCGATCAGG ATGATCTGGA CGAAGAGCAT CAGGGGCTCG CGCCAGCCGA ACTGTTCGCC     1740
AGGCTCAAGG CGAGCATGCC CGACGGCGAG GATCTCGTCG TGACCCATGG CGATGCCTGC     1800
TTGCCGAATA TCATGGTGGA AAATGGCCGC TTTTCTGGAT TCATCGACTG TGGCCGGCTG     1860
GGTGTGGCGG ACCGCTATCA GGACATAGCG TTGGCTACCC GTGATATTGC TGAAGAGCTT     1920
GGCGGCGAAT GGGCTGACCG CTTCCTCGTG CTTTACGGTA TCGCCGCTCC CGATTCGCAG     1980
CGCATCGCCT TCTATCGCCT TCTTGACGAG TTCTTCTGAG CGGGACTCTG GGGTTCGAAA     2040
TGACCGACCA AGCGACGCCC AACCTGCCAT CACGAGATTT CGATTCCACC GCCGCCTTCT     2100
ATGAAAGGTT GGGCTTCGGA ATCGTTTTCC GGGACGCCGG CTGGATGATC CTCCAGCGCG     2160
GGGATCTCAT GCTGGAGTTC TTCGCCCACC CTAGGGGGAG GCTAACTGAA ACACGGAAGG     2220
AGACAATACC GGAAGGAACC CGCGCTATGA CGGCAATAAA AAGACAGAAT AAAACGCACG     2280
GTGTTGGGTC GTTTGTTCAT AAACGCGGGG TTCGGTCCCA GGGCTGGCAC TCTGTCGATA     2340
CCCCACCGAG ACCCCATTGG GGCCAATACG CCCGCGTTTC TTCCTTTTCC CCACCCCACC     2400
CCCCAAGTTC GGGTGAAGGC CCAGGGCTCG CAGCCAACGT CGGGGCGGCA GGCCCTGCCA     2460
TAGCCTCAGG TTACTCATAT ATACTTTAGA TTGATTTAAA ACTTCATTTT TAATTTAAAA     2520
GGATCTAGGT GAAGATCCTT TTTGATAATC TCATGACCAA AATCCCTTAA CGTGAGTTTT     2580
CGTTCCACTG AGCGTCAGAC CCCGTAGAAA AGATCAAAGG ATCTTCTTGA GATCCTTTTT     2640
TTCTGCGCGT AATCTGCTGC TTGCAAACAA AAAAACCACC GCTACCAGCG GTGGTTTGTT     2700
TGCCGGATCA AGAGCTACCA ACTCTTTTTC CGAAGGTAAC TGGCTTCAGC AGAGCGCAGA     2760
TACCAAATAC TGTCCTTCTA GTGTAGCCGT AGTTAGGCCA CCACTTCAAG AACTCTGTAG     2820
CACCGCCTAC ATACCTCGCT CTGCTAATCC TGTTACCAGT GGCTGCTGCC AGTGGCGATA     2880
AGTCGTGTCT TACCGGGTTG GACTCAAGAC GATAGTTACC GGATAAGGCG CAGCGGTCGG     2940
GCTGAACGGG GGGTTCGTGC ACACAGCCCA GCTTGGAGCG AACGACCTAC ACCGAACTGA     3000
GATACCTACA GCGTGAGCTA TGAGAAAGCG CCACGCTTCC CGAAGGGAGA AAGGCGGACA     3060
GGTATCCGGT AAGCGGCAGG GTCGGAACAG GAGAGCGCAC GAGGGAGCTT CCAGGGGGAA     3120
ACGCCTGGTA TCTTTATAGT CCTGTCGGGT TTCGCCACCT CTGACTTGAG CGTCGATTTT     3180
TGTGATGCTC GTCAGGGGGG CGGAGCCTAT GGAAAAACGC CAGCAACGCG GCCTTTTTAC     3240
GGTTCCTGGC CTTTTGCTGG CCTTTTGCTC ACATGTTCTT TCCTGCGTTA TCCCCTGATT     3300
CTGTGGATAA CCGTATTACC GCCATGCATT AGTTATTAAT AGTAATCAAT TACGGGGTCA     3360
TTAGTTCATA GCCCATATAT GGAGTTCCGC GTTACATAAC TTACGGTAAA TGGCCCGCCT     3420
GGCTGACCGC CCAACGACCC CCGCCCATTG ACGTCAATAA TGACGTATGT TCCCATAGTA     3480
ACGCCAATAG GGACTTTCCA TTGACGTCAA TGGGTGGAGT ATTTACGGTA AACTGCCCAC     3540
TTGGCAGTAC ATCAAGTGTA TCATATGCCA AGTACGCCCC CTATTGACGT CAATGACGGT     3600
AAATGGCCCG CCTGGCATTA TGCCCAGTAC ATGACCTTAT GGGACTTTCC TACTTGGCAG     3660
TACATCTACG TATTAGTCAT CGCTATTACC ATGGTGATGC GGTTTTGGCA GTACATCAAT     3720
GGGCGTGGAT AGCGGTTTGA CTCACGGGGA TTTCCAAGTC TCCACCCCAT TGACGTCAAT     3780
GGGAGTTTGT TTTGGCACCA AAATCAACGG GACTTTCCAA AATGTCGTAA CAACTCCGCC     3840
CCATTGACGC AAATGGGCGG TAGGCGTGTA CGGTGGGAGG TCTATATAAG CAGAGCTGGT     3900
TTAGTGAACC GTCAGATCCG CTAGCGCTAC CGGTCGCCAC CATGGTGAGC AAGGGCGAGG     3960
AGCTGTTCAC CGGGGTGGTG CCCATCCTGG TCGAGCTGGA CGGCGACGTA AACGGCCACA     4020
AGTTCAGCGT GTCCGGCGAG GGCGAGGGCG ATGCCACCTA CGGCAAGCTG ACCCTGAAGT     4080
TCATCTGCAC CACCGGCAAG CTGCCCGTGC CCTGGCCCAC CCTCGTGACC ACCCTGACCT     4140
ACGGCGTGCA GTGCTTCAGC CGCTACCCCG ACCACATGAA GCAGCACGAC TTCTTCAAGT     4200
CCGCCATGCC CGAAGGCTAC GTCCAGGAGC GCACCATCTT CTTCAAGGAC GACGGCAACT     4260
ACAAGACCCG CGCCGAGGTG AAGTTCGAGG GCGACACCCT GGTGAACCGC ATCGAGCTGA     4320
AGGGCATCGA CTTCAAGGAG GACGGCAACA TCCTGGGGCA CAAGCTGGAG TACAACTACA     4380
ACAGCCACAA CGTCTATATC ATGGCCGACA AGCAGAAGAA CGGCATCAAG GTGAACTTCA     4440
AGATCCGCCA CAACATCGAG GACGGCAGCG TGCAGCTCGC CGACCACTAC CAGCAGAACA     4500
CCCCCATCGG CGACGGCCCC GTGCTGCTGC CCGACAACCA CTACCTGAGC ACCCAGTCCG     4560
CCCTGAGCAA AGACCCCAAC GAGAAGCGCG ATCACATGGT CCTGCTGGAG TTCGTGACCG     4620
CCGCCGGGAT CACTCTCGGC ATGGACGAGC TGTACAAGTC CGGACTCAGA TCTCGAGCTC     4680
AAGCTTCGAA TTCTGCAGTC GACGGTACCG CGGGCCCGGG ATCCAAGTAT CCAGATATTG     4740
CCTCACCCAC ATTTCGAAGG TTGTTTGGTG CCAAGGCAGG TGGCAAATCT GCCTCTGCAC     4800
CTAATACTGA GGGTGTGAAA TCTTCCTCAG TAATGCCCAG CCCTAGTACC ACATTAGCGC     4860
GGCAAGGCAG TCTGGAGTCA CCGTCGTCCG GTACGGGCAG CATGGGCAGT GCTGGTGGGC     4920
TAAGCGGCAG CAGCAGCCCT CTCTTCAATA AACCCTCAGA CTTAACTACA GATGTTATAA     4980
GCTTAAGTCA CTCGTTGGCC TCCAGCCCAG CATCGGTTCA CTCTTTCACA TCAGGTGGTC     5040
TCGTGTGGGC TGCCAATATG AGCAGTTCCT CTGCAGGCAG CAAGGATACT CCGAGCTACC     5100
AGTCCATGAC TAGCCTCCAC ACGAGCTCTG AGTCCATTGA CCTCCCCCTC AGCCATCATG     5160
GCTCCTTGTC TGGACTGACC ACAGGCACTC ACGAGGTCCA GAGCCTGCTC ATGAGAACGG     5220
GTAGTGTGAG ATCTACTCTC TCAGAAAGCA TGCAGCTTGA CAGAAATACA CTACCCAAAA     5280
AGGGACTAAG ATATACCCCA TCATCTCGGC AGGCCAACCA AGAAGAGGGC AAAGAGTGGT     5340
TGCGTTCTCA TTCTACTGGA GGGCTTCAGG ACACTGGCAA CCAGTCACCT CTGGTTTCCC     5400
```

Figure 7b (CONTINUED 2)

```
CTTCTGCCAT GTCATCTTCT GCAGCTGGAA AATACCACTT TTCTAACTTG GTGAGCCCAA   5460
CAAATTTGTC TCAATTTAAC CTTCCCGGGC CCAGCATGAT GCGCTCAAAC AGCATCCCAG   5520
CCCAAGACTC TTCCTTCGAT CTCTATGATG ACTCCCAGCT TTGTGGGAGT GCCACTTCTC   5580
TGGAGGAAAG ACCTCGTGCC ATCAGTCATT CGGGCTCATT CAGAGACAGC ATGGAAGAAG   5640
TTCATGGCTC TTCATTATCA CTGGTGTCCA GCACTTCTTC TCTTTACTCT ACAGCTGAAG   5700
AAAAGGCTCA TTCAGAGCAA ATCCATAAAC TGCGGAGAGA GCTGGTTGCA TCACAAGAAA   5760
AAGTTGCTAC CCTCACATCT CAGCTTTCAG CAAATGCTCA CCTTGTAGCA GCTTTTGAAA   5820
AGAGCTTAGG GAATATGACT GGCCGATTGC AAAGTCTAAC TATGACAGCG GAACAAAAGG   5880
AATCTGAACT TATAGAACTA AGAGAAACCA TTGAAATGCT GAAGGCTCAG AATTCTGCTG   5940
CCCAGGCGGC TATTCAGGGA GCACTGAATG GTCCAGACCA TCCTCCCAAA GATCTTCGCA   6000
TCAGAAGACA GCATTCCTCT GAAAGTGTTT CTAGTATCAA CAGTGCCACA AGCCATTCCA   6060
GTATTGGCAG TGGTAATGAT GCCGACTCCA AGAAGAAGAA AAAGAAAAAC TGGGTGAACT   6120
CTAGAGGAAG TGAGCTGAGA AGTTCTTTCA AACAAGCCTT TGGGAAGAAA AAGTCCACCA   6180
AGCCTCCTTC ATCACATTCT GACATTGAAG AGCTTACTGA TTCATCCCTT CCGGCATCCC   6240
CCAAGTTACC CCATAATGCT GGTGACTGTG GCTCAGCATC CATGAAGCCC TCACAATCTG   6300
CTTCAGCGAT CTGTGAATGC ACAGAAGCTG AGGCAGAGAT AATTCTGCAG CTGAAGAGCG   6360
AGCTCAGAGA AAAGGAATTA AAATTAACGG ATATTCGGCT GGAGGCCCTC AGCTCTGCTC   6420
ATCATCTTGA TCAGATCCGG GAAGCCATGA ACCGGATGCA GAATGAAATT GAAATACTGA   6480
AAGCTGAAAA TGACCGGTTG AAGGCAGAAA CTGGTAACAC AGCTAAGCCT ACTCGGCCAC   6540
CGTCAGAATC CTCAAGCAGC ACCTCCTCTT CATCTTCCAG GCAGTCATTA GGACTTTCTC   6600
TAAACAATTT GAACATCACA GAGGCTGTTA GCTCAGATAT TTTGCTAGAT GATGCTGGTG   6660
ATGCAACTGG ACATAAAGAT GGCCGCAGTG TGAAAATTAT AGTCTCCATA AGCAAGGGCT   6720
ATGGTCGAGC AAAGGACCAA AAATCTCAGG CATATTTGAT AGGCTCCATT GGTGTTAGTG   6780
GAAAAACCAA GTGGGATGTC TTAGATGGTG TAATAAGACG TCTCTTTAAG GAATATGTAT   6840
TCCGAATTGA TACATCCACT AGCCTTGGTC TGAGCTCTGA CTGCATTGCT AGCTACTGTA   6900
TAGGAGACTT AATTAGATCC CATAACCTAG AAGTGCCTGA ATTGCTGCCT TGTGGATACC   6960
TTGTTGGAGA TAATAACATC ATCACTGTGA ACCTCAAAGG GGTAGAAGAA AATAGTTTGG   7020
ACAGTTTTGT TTTTGATACG CTGATTCCTA AACCAATTAC CCAAAGGTAC TTTAACTTGT   7080
TGATGGAGCA TCACAGAATT ATACTCTCAG GACCGAGTGG TACTGGAAAG ACCTATTTGG   7140
CAAACAAACT TGCTGAATAT GTAATAACCA AATCTGGAAG GAAAAAAACA GAGGATGCAA   7200
TTGCCACTTT TAATGTGGAC CACAAGTCAA GTAAGGAATT GCAACAATAT CTAGCTAACC   7260
TGGCTGAACA GTGCAGTGCT GATAATAATG GAGTGGAGCT CCCAGTTGTA ATAATTCTTG   7320
ATAATCTTCA TCATGTGGGC TCTCTGAGTG ATATCTTCAA TGGTTTTCTC AATTGTAAAT   7380
ACAACAAATG TCCATATATT ATTGGAACAA TGAATCAGGG AGTTTCTTCA TCACCAAATC   7440
TAGAGCTGCA TCACAATTTC AGGTGGGTAT TATGTGCAAA TCATACAGAA CCAGTGAAAG   7500
GCTTTTTAGG CAGATATCTT CGAAGAAAAC TCATAGAGAT AGAAATTGAA AGGAACATTC   7560
GCAATAATGA CCTAGTCAAA ATTATAGATT GGATTCCGAA GACGTGGCAT CATCTCAACA   7620
GTTTTTTGGA AACACACAGT TCTTCTGACG TTACCATTGG TCCCCGACTA TTCCTTCCTT   7680
GCCCCATGGA TGTAGAAGGT TCTAGAGTAT GGTTCATGGA TCTCTGGAAC TATTCTTTAG   7740
TACCTTATAT TCTGGAGGCA GTGAGAGAGG GTCTTCAGAT GTATGGGAAA CGCACACCAT   7800
GGGAAGATCC TTCAAAGTGG GTGCTTGACA CATATCCATG GAGCTCAGCA ACTCTGCCTC   7860
AGGAGAGCCC AGCCTTACTT CAGCTGCGAC CAGAAGATGT TGGGTATGAA AGCTGCACAT   7920
CCACTAAGGA AGCCACAACC TCAAAGCACA TTCCGCAAAC TGACACAGAA GGAGATCCCC   7980
TGATGAATAT GCTAATGAAA CTCCAAGAAG CAGCCAATTA CTCAAGCACA CAAAGCTGCG   8040
ACAGCGAAAG CACCAGCCAC CATGAAGACA TTTTGGATTC ATCTCTTGAA TCTACCCTCT   8100
AGAGGGTGAA AGCCGAAATC CAGCACACTG GCGGCCGTTA                         8140
//
```

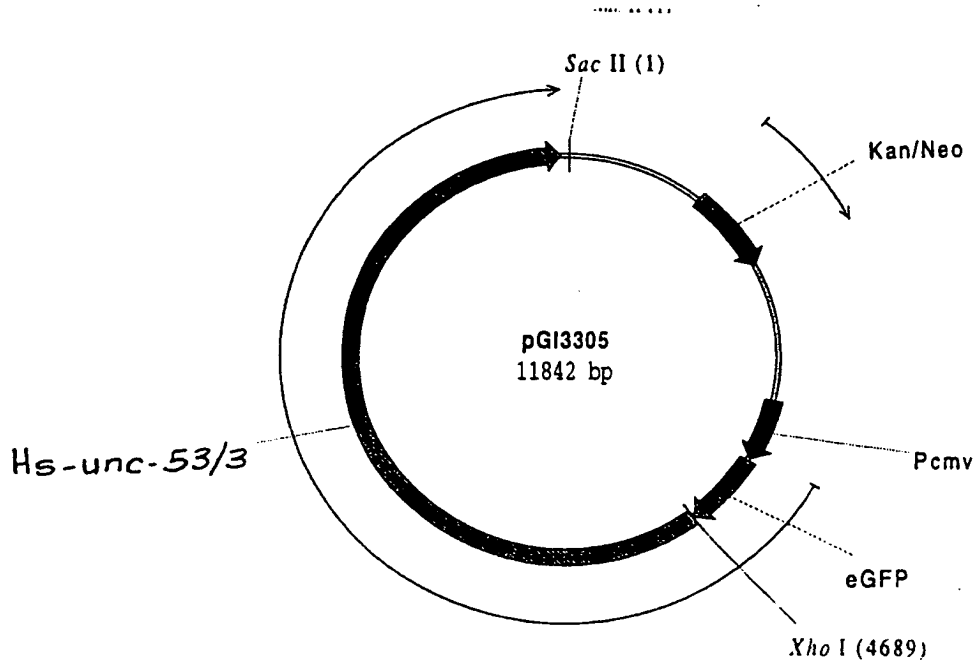
Legend: pGI3303 was obtained by inserting the 3421 bp BamHI/SpeI fragment of the Hs-Unc53/3GLd22_PCR2.1_D02 in a BamHI/XbaI opened pEGFPc1 vector (Clontech Inc.). This plasmid encodes an eGFP protein in fusion with the C-terminal half of Hs-unc-53/3 (last 1128 AA). Arrows indicate the ORFs.

Figure 7c: Illustration of the AA sequence of GFP::C-terminal Hs-unc-53/3 fragment(insert of pGI3303)

MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGV
QCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNIL
GHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSA
LSKDPNEKRDHMVLLEFVTAAGITLGMDELYKSGLRSRAQASNSAVDGTAGPGSKYPDIASPTFRRLFG
AKAGGKSASAPNTEGVKSSSVMPSPSTTLARQGSLESPSSGTGSMGSAGGLSGSSSPLFNKPSDLTTDV
ISLSHSLASSPASVHSFTSGGLVWAANMSSSSAGSKDTPSYQSMTSLHTSSESIDLPLSHHGSLSGLTT
GTHEVQSLLMRTGSVRSTLSESMQLDRNTLPKKGLRYTPSSRQANQEEGKEWLRSHSTGGLQDTGNQSP
LVSPSAMSSSAAGKYHFSNLVSPTNLSQFNLPGPSMMRSNSIPAQDSSFDLYDDSQLCGSATSLEERPR
AISHSGSFRDSMEEVHGSSLSLVSSTSSLYSTAEEKAHSEQIHKLRRELVASQEKVATLTSQLSANAHL
VAAFEKSLGNMTGRLQSLTMTAEQKESELIELRETIEMLKAQNSAAQAAIQGALNGPDHPPKDLRIRRQ
HSSESVSSINSATSHSSIGSGNDADSKKKKKKNWVNSRGSELRSSFKQAFGKKKSTKPPSSHSDIEELT
DSSLPASPKLPHNAGDCGSASMKPSQSASAICECTEAEAEIILQLKSELREKELKLTDIRLEALSSAHH
LDQIREAMNRMQNEIEILKAENDRLKAETGNTAKPTRPPSESSSSTSSSSSRQSLGLSLNNLNITEAVS
SDILLDDAGDATGHKDGRSVKIIVSISKGYGRAKDQKSQAYLIGSIGVSGKTKWDVLDGVIRRLFKEYV
FRIDTSTSLGLSSDCIASYCIGDLIRSHNLEVPELLPCGYLVGDNNIITVNLKGVEENSLDSFVFDTLI
PKPITQRYFNLLMEHHRIILSGPSGTGKTYLANKLAEYVITKSGRKKTEDAIATFNVDHKSSKELQQYL
ANLAEQCSADNNGVELPVVIILDNLHHVGSLSDIFNGFLNCKYNKCPYIIGTMNQGVSSSPNLELHHNF
RWVLCANHTEPVKGFLGRYLRRKLIEIEIERNIRNNDLVKIIDWIPKTWHHLNSFLETHSSSDVTIGPR
LFLPCPMDVEGSRVWFMDLWNYSLVPYILEAVREGLQMYGKRTPWEDPSKWVLDTYPWSSATLPQESPA
LLQLRPEDVGYESCTSTKEATTSKHIPQTDTEGDPLMNMLMKLQEAANYSSTQSCDSESTSHHEDILDS
SLESTL

Legend: Single underlined AA sequence represents eGFP. Double underlined AA sequence represents the C-terminal fragment of Hs-unc-53/3

Figure 7d: Illustration of the plasmid map and the nucleotide sequence of the pGI3305 expression vector (full length Hs-unc-53/3 in fusion with GFP)




```
ID   pGI3305                    circular DNA; 11842 BP
FT   CDS             1245..2039
FT                   /vntifkey="4"
FT                   /label=Kan/Neo
FT   CDS             3895..10983
FT                   /vntifkey="4"
FT                   /label=hHs-unc-53/3\(full\length)
FT   CDS             3962..4678
FT                   /vntifkey="4"
FT                   /label=eGFP
FT   promoter        3350..3938
FT                   /vntifkey="29"
FT                   /label=Pcmv
SQ   SEQUENCE   11842 BP;
     GGGCCCGGGA TCCACCGGAT CTAGATAACT GATCATAATC AGCCATACCA CATTTGTAGA      60
     GGTTTTACTT GCTTTAAAAA ACCTCCCACA CCTCCCCCTG AACCTGAAAC ATAAAATGAA     120
     TGCAATTGTT GTTGTTAACT TGTTTATTGC AGCTTATAAT GGTTACAAAT AAAGCAATAG     180
     CATCACAAAT TTCACAAATA AAGCATTTTT TTCACTGCAT TCTAGTTGTG GTTTGTCCAA     240
     ACTCATCAAT GTATCTTAAC GCGTAAATTG TAAGCGTTAA TATTTTGTTA AAATTCGCGT     300
     TAAATTTTTG TTAAATCAGC TCATTTTTTA ACCAATAGGC CGAAATCGGC AAAATCCCTT     360
     ATAAATCAAA AGAATAGACC GAGATAGGGT TGAGTGTTGT TCCAGTTTGG AACAAGAGTC     420
     CACTATTAAA GAACGTGGAC TCCAACGTCA AAGGGCGAAA AACCGTCTAT CAGGGCGATG     480
     GCCCACTACG TGAACCATCA CCCTAATCAA GTTTTTTGGG GTCGAGGTGC CGTAAAGCAC     540
     TAAATCGGAA CCCTAAAGGG AGCCCCCGAT TTAGAGCTTG ACGGGGAAAG CCGGCGAACG     600
     TGGCGAGAAA GGAAGGGAAG AAAGCGAAAG GAGCGGGCGC TAGGGCGCTG GCAAGTGTAG     660
     CGGTCACGCT GCGCGTAACC ACCACACCCG CCGCGCTTAA TGCGCCGCTA CAGGGCGCGT     720
     CAGGTGGCAC TTTTCGGGGA AATGTGCGCG GAACCCCTAT TTGTTTATTT TTCTAAATAC     780
     ATTCAAATAT GTATCCGCTC ATGAGACAAT AACCCTGATA AATGCTTCAA TAATATTGAA     840
     AAAGGAAGAG TCCTGAGGCG GAAAGAACCA GCTGTGGAAT GTGTGTCAGT TAGGGTGTGG     900
     AAAGTCCCCA GGCTCCCCAG CAGGCAGAAG TATGCAAAGC ATGCATCTCA ATTAGTCAGC     960
     AACCAGGTGT GGAAAGTCCC CAGGCTCCCC AGCAGGCAGA AGTATGCAAA GCATGCATCT    1020
     CAATTAGTCA GCAACCATAG TCCCGCCCCT AACTCCGCCC ATCCCGCCCC TAACTCCGCC    1080
     CAGTTCCGCC CATTCTCCGC CCCATGGCTG ACTAATTTTT TTTATTTATG CAGAGGCCGA    1140
```

Figure 7d (CONTINUED 1)

```
GGCCGCCTCG GCCTCTGAGC TATTCCAGAA GTAGTGAGGA GGCTTTTTTG GAGGCCTAGG  1200
CTTTTGCAAA GATCGATCAA GAGACAGGAT GAGGATCGTT TCGCATGATT GAACAAGATG  1260
GATTGCACGC AGGTTCTCCG GCCGCTTGGG TGGAGAGGCT ATTCGGCTAT GACTGGGCAC  1320
AACAGACAAT CGGCTGCTCT GATGCCGCCG TGTTCCGGCT GTCAGCGCAG GGGCGCCCGG  1380
TTCTTTTTGT CAAGACCGAC CTGTCCGGTG CCCTGAATGA ACTGCAAGAC GAGGCAGCGC  1440
GGCTATCGTG GCTGGCCACG ACGGGCGTTC CTTGCGCAGC TGTGCTCGAC GTTGTCACTG  1500
AAGCGGGAAG GGACTGGCTG CTATTGGGCG AAGTGCCGGG GCAGGATCTC CTGTCATCTC  1560
ACCTTGCTCC TGCCGAGAAA GTATCCATCA TGGCTGATGC AATGCGGCGG CTGCATACGC  1620
TTGATCCGGC TACCTGCCCA TTCGACCACC AAGCGAAACA TCGCATCGAG CGAGCACGTA  1680
CTCGGATGGA AGCCGGTCTT GTCGATCAGG ATGATCTGGA CGAAGAGCAT CAGGGGCTCG  1740
CGCCAGCCGA ACTGTTCGCC AGGCTCAAGG CGAGCATGCC CGACGGCGAG GATCTCGTCG  1800
TGACCCATGG CGATGCCTGC TTGCCGAATA TCATGGTGGA AAATGGCCGC TTTTCTGGAT  1860
TCATCGACTG TGGCCGGCTG GGTGTGGCGG ACCGCTATCA GGACATAGCG TTGGCTACCC  1920
GTGATATTGC TGAAGAGCTT GGCGGCGAAT GGGCTGACCG CTTCCTCGTG CTTTACGGTA  1980
TCGCCGCTCC CGATTCGCAG CGCATCGCCT TCTATCGCCT TCTTGACGAG TTCTTCTGAG  2040
CGGGACTCTG GGGTTCGAAA TGACCGACCA AGCGACGCCC AACCTGCCAT CACGAGATTT  2100
CGATTCCACC GCCGCCTTCT ATGAAAGGTT GGGCTTCGGA ATCGTTTTCC GGGACGCCGG  2160
CTGGATGATC CTCCAGCGCG GGGATCTCAT GCTGGAGTTC TTCGCCCACC CTAGGGGGAG  2220
GCTAACTGAA ACACGGAAGG AGACAATACC GGAAGGAACC CGCGCTATGA CGGCAATAAA  2280
AAGACAGAAT AAAACGCACG GTGTTGGGTC GTTTGTTCAT AAACGCGGGG TTCGGTCCCA  2340
GGGCTGGCAC TCTGTCGATA CCCCACCGAG ACCCCATTGG GGCCAATACG CCCGCGTTTC  2400
TTCCTTTTCC CCACCCCACC CCCCAAGTTC GGGTGAAGGC CCAGGGCTCG CAGCCAACGT  2460
CGGGGCGGCA GGCCCTGCCA TAGCCTCAGG TTACTCATAT ATACTTTAGA TTGATTTAAA  2520
ACTTCATTTT TAATTTAAAA GGATCTAGGT GAAGATCCTT TTTGATAATC TCATGACCAA  2580
AATCCCTTAA CGTGAGTTTT CGTTCCACTG AGCGTCAGAC CCCGTAGAAA AGATCAAAGG  2640
ATCTTCTTGA GATCCTTTTT TTCTGCGCGT AATCTGCTGC TTGCAAACAA AAAAACCACC  2700
GCTACCAGCG GTGGTTTGTT TGCCGGATCA AGAGCTACCA ACTCTTTTTC CGAAGGTAAC  2760
TGGCTTCAGC AGAGCGCAGA TACCAAATAC TGTCCTTCTA GTGTAGCCGT AGTTAGGCCA  2820
CCACTTCAAG AACTCTGTAG CACCGCCTAC ATACCTCGCT CTGCTAATCC TGTTACCAGT  2880
GGCTGCTGCC AGTGGCGATA AGTCGTGTCT TACCGGGTTG GACTCAAGAC GATAGTTACC  2940
GGATAAGGCG CAGCGGTCGG GCTGAACGGG GGGTTCGTGC ACACAGCCCA GCTTGGAGCG  3000
AACGACCTAC ACCGAACTGA GATACCTACA GCGTGAGCTA TGAGAAAGCG CCACGCTTCC  3060
CGAAGGGAGA AAGGCGGACA GGTATCCGGT AAGCGGCAGG GTCGGAACAG GAGAGCGCAC  3120
GAGGGAGCTT CCAGGGGGAA ACGCCTGGTA TCTTTATAGT CCTGTCGGGT TTCGCCACCT  3180
CTGACTTGAG CGTCGATTTT TGTGATGCTC GTCAGGGGGG CGGAGCCTAT GGAAAAACGC  3240
CAGCAACGCG GCCTTTTTAC GGTTCCTGGC CTTTTGCTGG CCTTTTGCTC ACATGTTCTT  3300
TCCTGCGTTA TCCCCTGATT CTGTGGATAA CCGTATTACC GCCATGCATT AGTTATTAAT  3360
AGTAATCAAT TACGGGGTCA TTAGTTCATA GCCCATATAT GGAGTTCCGC GTTACATAAC  3420
TTACGGTAAA TGGCCCGCCT GGCTGACCGC CCAACGACCC CCGCCCATTG ACGTCAATAA  3480
TGACGTATGT TCCCATAGTA ACGCCAATAG GGACTTTCCA TTGACGTCAA TGGGTGGAGT  3540
ATTTACGGTA AACTGCCCAC TTGGCAGTAC ATCAAGTGTA TCATATGCCA AGTACGCCCC  3600
CTATTGACGT CAATGACGGT AAATGGCCCG CCTGGCATTA TGCCCAGTAC ATGACCTTAT  3660
GGGACTTTCC TACTTGGCAG TACATCTACG TATTAGTCAT CGCTATTACC ATGGTGATGC  3720
GGTTTTGGCA GTACATCAAT GGGCGTGGAT AGCGGTTTGA CTCACGGGGA TTTCCAAGTC  3780
TCCACCCCAT TGACGTCAAT GGGAGTTTGT TTTGGCACCA AAATCAACGG GACTTTCCAA  3840
AATGTCGTAA CAACTCCGCC CCATTGACGC AAATGGGCGG TAGGCGTGTA CGGTGGGAGG  3900
TCTATATAAG CAGAGCTGGT TTAGTGAACC GTCAGATCCG CTAGCGCTAC CGGTCGCCAC  3960
CATGGTGAGC AAGGGCGAGG AGCTGTTCAC CGGGGTGGTG CCCATCCTGG TCGAGCTGGA  4020
CGGCGACGTA AACGGCCACA AGTTCAGCGT GTCCGGCGAG GGCGAGGGCG ATGCCACCTA  4080
CGGCAAGCTG ACCCTGAAGT TCATCTGCAC CACCGGCAAG CTGCCCGTGC CCTGGCCCAC  4140
CCTCGTGACC ACCCTGACCT ACGGCGTGCA GTGCTTCAGC CGCTACCCCG ACCACATGAA  4200
GCAGCACGAC TTCTTCAAGT CCGCCATGCC CGAAGGCTAC GTCCAGGAGC GCACCATCTT  4260
CTTCAAGGAC GACGGCAACT ACAAGACCCG CGCCGAGGTG AAGTTCGAGG GCGACACCCT  4320
GGTGAACCGC ATCGAGCTGA AGGGCATCGA CTTCAAGGAG GACGGCAACA TCCTGGGGCA  4380
CAAGCTGGAG TACAACTACA ACAGCCACAA CGTCTATATC ATGGCCGACA AGCAGAAGAA  4440
CGGCATCAAG GTGAACTTCA AGATCCGCCA CAACATCGAG GACGGCAGCG TGCAGCTCGC  4500
CGACCACTAC CAGCAGAACA CCCCCATCGG CGACGGCCCC GTGCTGCTGC CCGACAACCA  4560
CTACCTGAGC ACCCAGTCCG CCCTGAGCAA AGACCCCAAC GAGAAGCGCG ATCACATGGT  4620
CCTGCTGGAG TTCGTGACCG CCGCCGGGAT CACTCTCGGC ATGGACGAGC TGTACAAGTA  4680
CTCAGATCTC GAGCATATGC CTGTTCTTGG GGTTGCCTCA AAACTGAGGC AGCCAGCTGT  4740
TGGGTCAAAG CCTGTGCATA CTGCTCTTCC GATACCAAAT CTTGGCACTA CTGGGTCACA  4800
GCACTGTTCT TCAAGACCTT TGGAACTTGC TGAAACAGAG AGCTCCATGC TTTCTTGTCA  4860
GCTTGCGTTA AAATCAACCT GTGAATTTGG AGAGAAGAAA CCCCTCCAAG GAAAAGCCAA  4920
GGAGAAAGAA GACAGCAAGA TTTACACTGA CTGGGCCAAC CACTACCTAG CAAAATCAGG  4980
CCACAAGCGG CTGATCAAGG ACTTGCAACA AGACATTGCA GATGGAGTAC TCCTAGCAGA  5040
AATCATCCAG ATTATTGCAA ATGAAAAAGT TGAAGATATC AATGGATGTC CTAGAAGTCA  5100
GTCTCAGATG ATTGAAAATG TTGATGTCTG CCTTAGTTTT CTAGCAGCCA GAGGGGTAAA  5160
TGTTCAAGGT CTATCTGCTG AAGAAATAAG AAATGGAAAC TTAAAAGCCA TTCTAGGGCT  5220
GTTTTTCAGT TTATCTCGCT ACAAGCAGCA ACAACACCAT CAACAACAGT ACTATCAGTC  5280
CTTGGTGGAA CTTCAGCAGC GAGTTACTCA CGCTTCCCCT CCATCGGAAG CCAGCCAGGC  5340
CAAAACCCAG CAAGATATGC AGTCCAGTCT GGCAGCCAGA TATGCAACTC AGTCTAATCA  5400
CAGTGGAATT GCAACCAGTC AAAAAAAGCC TACTAGGCTT CCAGGGCCCT CTAGGGTGCC  5460
TGCTGCAGGA AGCAGCAGCA AGGTCCAGGG AGCCTCTAAT TTAAATAGGA GAAGTCAGAG  5520
CTTTAACAGC ATTGACAAAA ACAAGCCTCC AAATTATGCA AATGGAAACG AAAAAGATTC  5580
CTCCAAAGGA CCTCAATCGT CTTCAGGTGT AAATGGTAAC GTGCAGCCTC CCAGTACTGC  5640
TGGGCAGCCT CCTGCCTCTG CCATCCCTTC TCCAAGTGCC AGCAAGCCCT GGCGCAGCAA  5700
```

Figure 7d (CONTINUED 2)

```
GTCCATGAAT GTCAAACACA GTGCCACCTC CACCATGTTG ACTGTAAAGC AGTCAAGTAC      5760
AGCCACCTCC CCCACACCAT CTTCAGACAG ACTGAAGCCA CCTGTCTCAG AAGGGGTCAA      5820
AACTGCTCCC TCAGGACAGA AATCCATGCT TGAGAAATTC AAGCTAGTCA ATGCCCGGAC      5880
TGCTTTACGC CCCCCGCAGC CTCCCAGTTC AGGACCTAGT GATGGTGGGA AGGATGATGA      5940
TGCCTTTTCT GAATCTGGTG AAATGGAAGG TTTTAACAGT GGTCTGAATA GTGGTGGCTC      6000
AACAAATAGC AGTCCCAAAG TGTCACCTAA GTTGGCCCCT CCAAAAGCTG GAAGCAAAAA      6060
TCTCAGCAAT AAAAAGTCTT TGCTACAGCC AAAGGAAAAA GAAGAAAAGA ACAGGGACAA      6120
AAATAAAGTT TGCACTGAAA AACCAGTCAA AGAAGAGAAG GATCAGGTGA CAGAGATGGC      6180
TCCAAAAAAG ACCTCCAAAA TTGCAAGCTT GATCCCTAAG GGCAGCAAGA CAACAGCAGC      6240
TAAGAAGGAA AGCTTAATTC CGTCTTCCAG TGGTATTCCA AAACCAGGCT CTAAAGTTCC      6300
AACAGTAAAG CAAACCATTT CACCTGGCAG CACAGCAAGC AAAGAGTCTG AGAAATTCAG      6360
GACTACCAAG GGGAGCCCTT CCCAGTCCTT ATCTAAGCCT ATAACCATGG AGAAAGCAAG      6420
TGCTTCTAGT TGTCCTGCCC CTTTGGAAGG AAGGGAAGCT GGCCAAGCTT CTCCTTCTGG      6480
TTCCTGTACC ATGACAGTGG CACAAAGCAG TGGGCAGAGC ACAGGAAATG GTGCTGTCCA      6540
ACTCCCTCAA CAGCAGCAAC ATAGCCACCC GAATACCGCG ACAGTGGCAC CATTCATTTA      6600
CAGGGCACAT TCAGAAAATG AAGGTACCGC TTTACCATCG GCTGACTCCT GTACCAGTCC      6660
TACAAAGATG GACTTATCAT ATAGTAAGAC TGCTAAGCAG TGCCTGGAGG AGATATCTGG      6720
TGAAGACCCT GAAACAAGAA GAATGAGAAC AGTTAAAAAC ATAGCAGACT TGAGGCAGAA      6780
TTTAGAAGAG ACTATGTCCA GTCTTCGTGG GACTCAGATA AGCCACAGCA CCCTGGAGAC      6840
AACATTTGAC AGCACTGTGA CAACAGAAGT TAATGGAAGG ACCATACCCA ACTTGACAAG      6900
TCGACCCACC CCCATGACCT GGAGGTTGGG CCAGGCATGT CCGCGACTTC AGGCGGGAGA      6960
TGCTCCCTCC CTGGGTGCTG GCTATCCTCG CAGTGGTACC AGTCGATTCA TCCACACAGA      7020
CCCCTCAAGG TTCATGTATA CCACGCCTCT CCGTCGAGCT GCTGTCTCTA GGCTGGGAAA      7080
CATGTCACAG ATTGACATGA GTGAGAAAGC AAGCAGTGAC CTGGACATGT CTTCTGAGGT      7140
CGATGTGGGT GGATATATGA GTGATGGTGA TATCCTTGGG AAAAGTCTCA GGACTGATGA      7200
CATCAACAGT GGGTACATGA CAGATGGAGG ACTTAACCTA TATACTAGAA GTCTGAACCG      7260
AATACCAGAC ACAGCAACTT CCCGGGACAT CATCCAGAGA GGGGTTCACG ATGTGACAGT      7320
GGATGCAGAC AGCTGGGATG ACAGCAGTTC AGTGAGCAGT GGTCTCAGTG ACACCCTTGA      7380
TAACATCAGC ACTGATGACC TGAACACCAC ATCCTCTGTC AGCTCTTACT CCAACATCAC      7440
CGTCCCCTCT AGGAAGAATA CTCAGCTGAG GACAGATTCA GAGAAACGCT CCACCACAGA      7500
CGAGACCTGG GATAGTCCTG AGGAACTGAA AAAACCAGAA NAAGATTTTG ACAGCCATGG      7560
GGATGCTGGT GGCAAGTGGA AGACTGTGTC CTCTGGACTT CCTGAAGACC CCGAGAAGGC      7620
AGGGCAGAAA GCTTCCCTGT CTGTTTCACA GACAGGTTCC TGGAGAAGAG GCATGTCTGC      7680
CCAAGGAGGG GCGCCATCTA GGCAGAAAGC TGGAACAAGT GCACTCAAAA CACCCGGGAA      7740
AACCGATGAT GCCAAAGCTT CTGAGAAAGG AAAAGCTCCC CTAAAAGGAT CATCTCTACA      7800
AAGATCTCCT TCAGATGCAG GAAAAAGCAG TGGAGATGAA GGGAAAAAGC CCCCCTCAGG      7860
CATTGGAAGA TCGACTGCCA CCAGCTCCTT TGGCTTTAAG AAACCAAGTG GAGTAGGGTC      7920
ATCTGCCATG ATCACCAGCA GTGGAGCAAC CATAACAAGT GGCTCTGCAA CACTGGGTAA      7980
AATTCCAAAA TCTGCTGCCA TTGGCGGGAA GTCAAATGCA GGGAGAAAAA CCAGTTTGGA      8040
CGGTTCACAG AATCAGGATG ATGTTGTGCT GCATGTTAGC TCAAAGACTA CCCTACAATA      8100
TCGCAGCTTG CCCCGCCCTT CAAAATCCAG CACCAGTGGC ATTCCTGGAC GAGGAGGCCA      8160
CAGATCCAGT ACCAGCAGTA TTGATTCCAA CGTCAGCAGC AAGTCTGCTG GGGCCACCAC      8220
CTCGAAACTG AGAGAACCAA CTAAAATTGG GTCAGGGCGC TCAAGTCCTG TCACCGTCAA      8280
CCAAACAGAC AAGGAAAAGG AAAAAGTAGC AGTCTCAGAT TCAGAAAGTG TTTCTTTGTC      8340
AGGTTCCCCC AAATCCAGCC CCACCTCTGC CAGCGCCTGT GGTGCACAAG GTCTCAGGCA      8400
GCCAGGATCC AAGTATCCAG ATATTGCCTC ACCCACATTT CGAAGGTTGT TTGGTGCCAA      8460
GGCAGGTGGC AAATCTGCCT CTGCACCTAA TACTGAGGGT GTGAAATCTT CCTCAGTAAT      8520
GCCCAGCCCT AGTACCACAT TAGCGCGGCA AGGCAGTCTG GAGTCACCGT CGTCCGGTAC      8580
GGGCAGCATG GGCAGTGCTG GTGGGCTAAG CGGCAGCAGC AGCCCTCTCT TCAATAAACC      8640
CTCAGACTTA ACTACAGATG TTATAAGCTT AAGTCACTCG TTGGCCTCCA GCCCAGCATC      8700
GGTTCACTCT TTCACATCAG GTGGTCTCGT GTGGGCTGCC AATATGAGCA GTTCCTCTGC      8760
AGGCAGCAAG GATACTCCGA GCTACCAGTC CATGACTAGC CTCCACACGA GCTCTGAGTC      8820
CATTGACCTC CCCCTCAGCC ATCATGGCTC CTTGTCTGGA CTGACCACAG GCACTCACGA      8880
GGTCCAGAGC CTGCTCATGA GAACGGGTAG TGTGAGATCT ACTCTCTCAG AAAGCATGCA      8940
GCTTGACAGA AATACACTAC CCAAAAAGGG ACTAAGATAT ACCCCATCAT CTCGGCAGGC      9000
CAACCAAGAA GAGGGCAAAG AGTGGTTGCG TTCTCATTCT ACTGGAGGGC TTCAGGACAC      9060
TGGCAACCAG TCACCTCTGG TTTCCCCTTC TGCCATGTCA TCTTCTGCAG CTGGAAAATA      9120
CCACTTTTCT AACTTGGTGA GCCCAACAAA TTTGTCTCAA TTTAACCTTC CCGGGCCCAG      9180
CATGATGCGC TCAAACAGCA TCCCAGCCCA AGACTCTTCC TTCGATCTCT ATGATGACTC      9240
CCAGCTTTGT GGGAGTGCCA CTTCTCTGGA GGAAAGACCT CGTGCCATCA GTCATTCGGG      9300
CTCATTCAGA GACAGCATGG AAGAAGTTCA TGGCTCTTCA TTATCACTGG TGTCCAGCAC      9360
TTCTTCTCTT TACTCTACAG CTGAAGAAAA GGCTCATTCA GAGCAAATCC ATAAACTGCG      9420
GAGAGAGCTG GTTGCATCAC AAGAAAAAGT TGCTACCCTC ACATCTCAGC TTTCAGCAAA      9480
TGCTCACCTT GTAGCAGCTT TTGAAAAGAG CTTAGGGAAT ATGACTGGCC GATTGCAAAG      9540
TCTAACTATG ACAGCGGAAC AAAAGGAATC TGAACTTATA GAACTAAGAG AAACCATTGA      9600
AATGCTGAAG GCTCAGAATT CTGCTGCCCA GGCGGCTATT CAGGGAGCAC TGAATGGTCC      9660
AGACCATCCT CCCAAAGATC TTCGCATCAG AAGACAGCAT TCCTCTGAAA GTGTTTCTAG      9720
TATCAACAGT GCCACAAGCC ATTCCAGTAT TGGCAGTGGT AATGATGCCG ACTCCAAGAA      9780
GAAGAAAAAG AAAAACTGGG TGAACTCTAG AGGAAGTGAG CTGAGAAGTT CTTTCAAACA      9840
AGCCTTTGGG AAGAAAAAGT CCACCAAGCC TCCTTCATCA CATTCTGACA TTGAAGAGCT      9900
TACTGATTCA TCCCTTCCGG CATCCCCCAA GTTACCCCAT AATGCTGGTG ACTGTGGCTC      9960
AGCATCCATG AAGCCCTCAC AATCTGCTTC AGCGATCTGT GAATGCACAG AAGCTGAGGC     10020
AGAGATAATT CTGCAGCTGA AGAGCGAGCT CAGAGAAAAG GAATTAAAAT TAACGGATAT     10080
TCGGCTGGAG GCCCTCAGCT CTGCTCATCA TCTTGATCAG ATCCGGGAAG CCATGAACCG     10140
GATGCAGAAT GAAATTGAAA TACTGAAAGC TGAAAATGAC CGGTTGAAGG CAGAAACTGG     10200
TAACACAGCT AAGCCTACTC GGCCACCGTC AGAATCCTCA AGCAGCACCT CCTCTTCATC     10260
```

Figure 7d (CONTINUED 3)

```
TTCCAGGCAG TCATTAGGAC TTTCTCTAAA CAATTTGAAC ATCACAGAGG CTGTTAGCTC    10320
AGATATTTTG CTAGATGATG CTGGTGATGC AACTGGACAT AAAGATGGCC GCAGTGTGAA    10380
AATTATAGTC TCCATAAGCA AGGGCTATGG TCGAGCAAAG GACCAAAAAT CTCAGGCATA    10440
TTTGATAGGC TCCATTGGTG TTAGTGGAAA AACCAAGTGG GATGTCTTAG ATGGTGTAAT    10500
AAGACGTCTC TTTAAGGAAT ATGTATTCCG AATTGATACA TCCACTAGCC TTGGTCTGAG    10560
CTCTGACTGC ATTGCTAGCT ACTGTATAGG AGACTTAATT AGATCCCATA ACCTAGAAGT    10620
GCCTGAATTG CTGCCTTGTG GATACCTTGT TGGAGATAAT AACATCATCA CTGTGAACCT    10680
CAAAGGGGTA GAAGAAAATA GTTTGGACAG TTTTGTTTTT GATACGCTGA TTCCTAAACC    10740
AATTACCCAA AGGTACTTTA ACTTGTTGAT GGAGCATCAC AGAATTATAC TCTCAGGACC    10800
GAGTGGTACT GGAAAGACCT ATTTGGCAAA CAAACTTGCT GAATATGTAA TAACCAAATC    10860
TGGAAGGAAA AAAACAGAGG ATGCAATTGC CACTTTTAAT GTGGACCACA AGTCAAGTAA    10920
GGAATTGCAA CAATATCTAG CTAACCTGGC TGAACAGTGC AGTGCTGATA ATAATGGAGT    10980
GGAGCTCCCA GTTGTAATAA TTCTTGATAA TCTTCATCAT GTGGGCTCTC TGAGTGATAT    11040
CTTCAATGGT TTTCTCAATT GTAAATACAA CAAATGTCCA TATATTATTG GAACAATGAA    11100
TCAGGGAGTT TCTTCATCAC CAAATCTAGA GCTGCATCAC AATTTCAGGT GGGTATTATG    11160
TGCAAATCAT ACAGAACCAG TGAAAGGCTT TTTAGGCAGA TATCTTCGAA GAAAACTCAT    11220
AGAGATAGAA ATTGAAAGGA ACATTCGCAA TAATGACCTA GTCAAAATTA TAGATTGGAT    11280
TCCGAAGACG TGGCATCATC TCAACAGTTT TTTGGAAACA CACAGTTCTT CTGACGTTAC    11340
CATTGGTCCC CGACTATTCC TTCCTTGCCC CATGGATGTA GAAGGTTCTA GAGTATGGTT    11400
CATGGATCTC TGGAACTATT CTTTAGTACC TTATATTCTG GAGGCAGTGA GAGAGGGTCT    11460
TCAGATGTAT GGGAAACGCA CACCATGGGA AGATCCTTCA AAGTGGGTGC TTGACACATA    11520
TCCATGGAGC TCAGCAACTC TGCCTCAGGA GAGCCCAGCC TTACTTCAGC TGCGACCAGA    11580
AGATGTTGGG TATGAAAGCT GCACATCCAC TAAGGAAGCC ACAACCTCAA AGCACATTCC    11640
GCAAACTGAC ACAGAAGGAG ATCCCCTGAT GAATATGCTA ATGAAACTCC AAGAAGCAGC    11700
CAATTACTCA AGCACACAAA GCTGCGACAG CGAAAGCACC AGCCACCATG AAGACATTTT    11760
GGATTCATCT CTTGAATCTA CCCTCTAGAG GGTGAAAGCC GAAATCCAGC ACACTGGCGG    11820
CCGTTACTAG TGGATCGGCC GC                                              11842
//
```

Legend: pGI3305 was obtained by inserting a 7148 bp XhoI/SacII fragment of the Hs-unc-53/3A1Ld22 clone in a XhoI/SacII opened pEGFPc3 vector (Clontech Inc.). This plasmid encodes an eGFP protein in fusion with the full length Hs-unc-53/3 (2363 AA). Arrows indicate the ORFs.

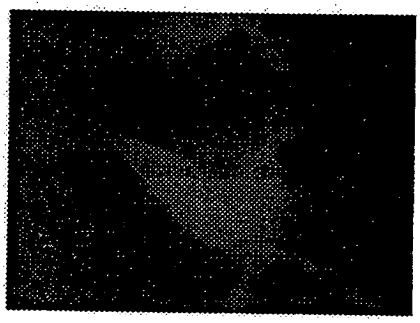
Figure 7e: Illustration of the AA sequence of GFP::Hs-unc-53/3 (insert of pGI3305)

MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGV
QCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNIL
GHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSA
LSKDPNEKRDHMVLLEFVTAAGITLGMDELYKYSDLEHMPVLGVASKLRQPAVGSKPVHTALPIPNLGT
TGSQHCSSRPLELAETESSMLSCQLALKSTCEFGEKKPLQGKAKEKEDSKIYTDWANHYLAKSGHKRLI
KDLQQDIADGVLLAEIIQIIANEKVEDINGCPRSQSQMIENVDVCLSFLAARGVNVQGLSAEEIRNGNL
KAILGLFFSLSRYKQQQHHQQQYYQSLVELQQRVTHASPPSEASQAKTQQDMQSSLAARYATQSNHSGI
ATSQKKPTRLPGPSRVPAAGSSSKVQGASNLNRRSQSFNSIDKNKPPNYANGNEKDSSKGPQSSSGVNG
NVQPPSTAGQPPASAIPSPSASKPWRSKSMNVKHSATSTMLTVKQSSTATSPTPSSDRLKPPVSEGVKT
APSGQKSMLEKFKLVNARTALRPPQPPSSGPSDGGKDDDAFSESGEMEGFNSGLNSGGSTNSSPKVSPK
LAPPKAGSKNLSNKKSLLQPKEKEEKNRDKNKVCTEKPVKEEKDQVTEMAPKKTSKIASLIPKGSKTTA
AKKESLIPSSSGIPKPGSKVPTVKQTISPGSTASKESEKFRTTKGSPSQSLSKPITMEKASASSCPAPL
EGREAGQASPSGSCTMTVAQSSGQSTGNGAVQLPQQQQHSHPNTATVAPFIYRAHSENEGTALPSADSC
TSPTKMDLSYSKTAKQCLEEISGEDPETRRMRTVKNIADLRQNLEETMSSLRGTQISHSTLETTFDSTV
TTEVNGRTIPNLTSRPTPMTWRLGQACPRLQAGDAPSLGAGYPRSGTSRFIHTDPSRFMYTTPLRRAAV
SRLGNMSQIDMSEKASSDLDMSSEVDVGGYMSDGDILGKSLRTDDINSGYMTDGGLNLYTRSLNRIPDT
ATSRDIIQRGVHDVTVDADSWDDSSSVSSGLSDTLDNISTDDLNTTSSVSSYSNITVPSRKNTQLRTDS
EKRSTTDETWDSPEELKKPEXDFDSHGDAGGKWKTVSSGLPEDPEKAGQKASLSVSQTGSWRRGMSAQG
GAPSRQKAGTSALKTPGKTDDAKASEKGKAPLKGSSLQRSPSDAGKSSGDEGKKPPSGIGRSTATSSFG
FKKPSGVGSSAMITSSGATITSGSATLGKIPKSAAIGGKSNAGRKTSLDGSQNQDDVVLHVSSKTTLQY
RSLPRPSKSSTSGIPGRGGHRSSTSSIDSNVSSKSAGATTSKLREPTKIGSGRSSPVTVNQTDKEKEKV
AVSDSESVSLSGSPKSSPTSASACGAQGLRQPGSKYPDIASPTFRRLFGAKAGGKSASAPNTEGVKSSS
VMPSPSTTLARQGSLESPSSGTGSMGSAGGLSGSSSPLFNKPSDLTTDVISLSHSLASSPASVHSFTSG
GLVWAANMSSSSAGSKDTPSYQSMTSLHTSSESIDLPLSHHGSLSGLTTGTHEVQSLLMRTGSVRSTLS
ESMQLDRNTLPKKGLRYTPSSRQANQEEGKEWLRSHSTGGLQDTGNQSPLVSPSAMSSSAAGKYHFSNL
VSPTNLSQFNLPGPSMMRSNSIPAQDSSFDLYDDSQLCGSATSLEERPRAISHSGSFRDSMEEVHGSSL
SLVSSTSSLYSTAEEKAHSEQIHKLRRELVASQEKVATLTSQLSANAHLVAAFEKSLGNMTGRLQSLTM
TAEQKESELIELRETIEMLKAQNSAAQAAIQGALNGPDHPPKDLRIRRQHSSESVSSINSATSHSSIGS
GNDADSKKKKKKNWVNSRGSELRSSFKQAFGKKKSTKPPSSHSDIEELTDSSLPASPKLPHNAGDCGSA
SMKPSQSASAICECTEAEAEIILQLKSELREKELKLTDIRLEALSSAHHLDQIREAMNRMQNEIEILKA
ENDRLKAETGNTAKPTRPPSESSSSTSSSSSRQSLGLSLNNLNITEAVSSDILLLDDAGDATGHKDGRSV
KIIVSISKGYGRAKDQKSQAYLIGSIGVSGKTKWDVLDGVIRRLFKEYVFRIDTSTSLGLSSDCIASYC
IGDLIRSHNLEVPELLPCGYLVGDNNIITVNLKGVEENSLDSFVFDTLIPKPITQRYFNLLMEHHRIIL
SGPSGTGKTYLANKLAEYVITKSGRKKTEDAIATFNVDHKSSKELQQYLANLAEQCSADNNGVELPVVI
ILDNLHHVGSLSDIFNGFLNCKYNKCPYIIGTMNQGVSSSPNLELHHNFRWVLCANHTEPVKGFLGRYL
RRKLIEIEIERNIRNNDLVKIIDWIPKTWHHLNSFLETHSSSDVTIGPRLFLPCPMDVEGSRVWFMDLW
NYSLVPYILEAVREGLQMYGKRTPWEDPSKWVLDTYPWSSATLPQESPALLQLRPEDVGYESCTSTKEA
TTSKHIPQTDTEGDPLMNMLMKLQEAANYSSTQSCDSESTSHHEDILDSSLESTL

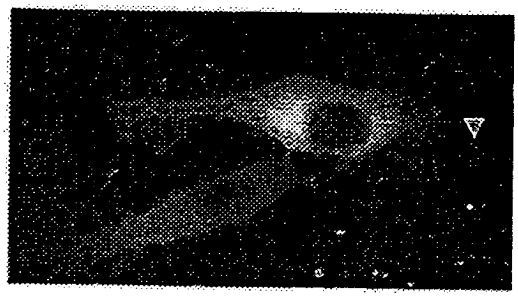
Legend: Single underlined AA sequence represents eGFP. Double underlined AA sequence represents full length Hs-unc-53/3.

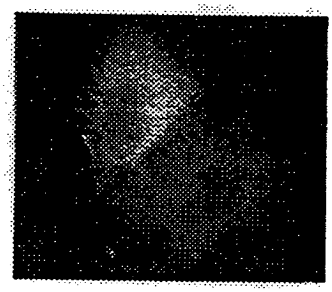
FIG. 8 Illustration of the filopodia and lamellipodia outgrowth of N4 mouse neuroblastoma cells transfected with pGI3303.

A:

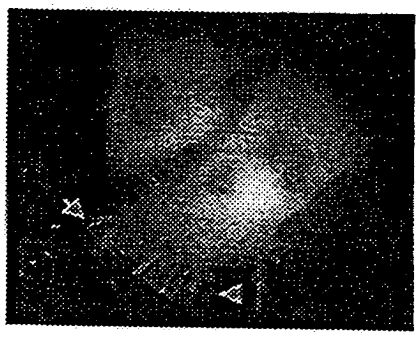


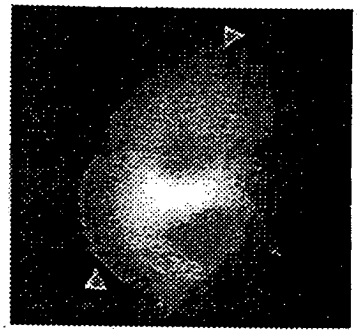
B:





C:

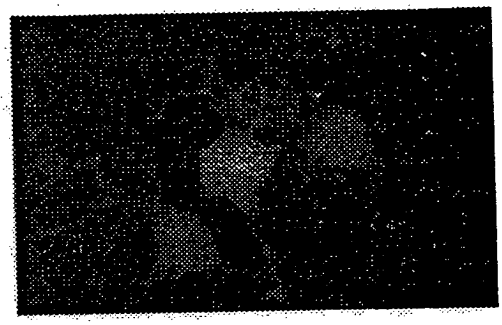




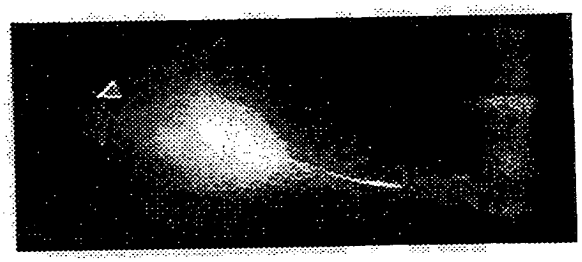
Legend: Fluorescence images of N4 cells transfected with pEGFP (A) compared to pGI3303 transfected cells (B and C). A: control (pEGFP) transfected cells. B: Illustration of filopodia outgrowth (arrowhead). C: Illustration of lamellipodia outgrowth (arrowhead). Notice the actin sheets at the edge of the cells.

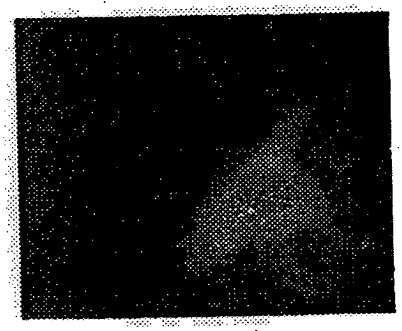
FIG. 9 Illustration of the co-localization of the GFP-Hs-unc-53/3 fusion protein with microtubules in N4 mouse neuroblastoma cells transfected with pGI3305

A:



B:





C:

Legend: Fluorescence images of N4 cells transfected with pEGFP (A) compared to pGI3305 transfected cells (B and C). A: control transfected cells. B: Illustration of co-localization of Hs-unc-53/3 with microtubuli. Notice the centrosome in the right picture (arrowhead) and enhanced filopodia outgrowth in the left picture (arrowhead). C: Illustration of the co-localization of Hs-unc-53/3 with(+)-end of microtubules (arrowhead).

Figure 11a: Illustration of the homology between Hs-unc-53/3 and a gene encoded (partially) by the Drosophila melanogaster BAC clone BACR48M05 (AC005719). Results of a TBLASTN search on the non-redundant database with Hs-unc-53/3 as query.

```
Query:          Hs-unc-53/3 (direct) 2120aa Length 2119 from:1 to = 2119
Sbjct:          gb|AC005719|AC005719 Drosophila melanogaster, chromosome 2R, region
                38A5-38B4, BAC clone
                BACR48M05, complete sequence [Drosophila melanogaster]
                Length = 188357


Score = 64.0 bits (153), Expect = 4e-08
Identities = 28/58 (48%), Positives = 41/58 (70%)


Query: 1      IYTDWANHYLAKSGHKRLIKDLQQDIADGVLLAEIIQIIANEKVEDINGCPRSQSQMI 58
                IYTDWAN+YL ++  KR + DL  D  DG+LLAE+I+ + + KV D+   P++Q QM+
Sbjct: 84874  IYTDWANYYLERAKSKRKVTDLSADCRDGLLLAEVIEAVTSFKVPDLVKKPKNQQQMV 84701


 Score = 39.9 bits (91), Expect = 0.77
 Identities = 22/55 (40%), Positives = 34/55 (61%)


Query: 48     NGCPRSQSQMIENVDVCLSFLAARGVN-VQGLSAEEIRNGNLKAILGLFFSLSRYK 102
               N C   Q    +NV+ CL  L ++ V  ++ ++  +I  G LKA+L LFF+LSR+K
Sbjct: 55621  NSCSLFQ---FDNVNSCLHVLRSQSVGGLENITTNDICAGRLKAVLALFFALSRFK 55463


 Score = 35.2 bits (79), Expect = 3.8
 Identities = 31/72 (43%), Positives = 45/72 (62%)


Query: 1266   LEERPRAISHSGSFRDSMEEVHGSSLSLVSSTSSLYSTAEEKAHSEQIHKLRRELVASQE 1325
               L+ R   + HS S       VHGS  SL+S  SSLY  AEE+  + +I +L+REL  +++
Sbjct: 13387  LKSRLMQLCHSVSV-----SVHGSAASLLSGGSSLYGNAEER-QAHEIRRLKRELQDARD 13226


Query: 1326   KVATLTSQLSAN 1337
               +V +L+SQLS N
Sbjct: 13225  QVLSLSSQLSTN 13190
```

Figure 11b: Illustration of an ORF encoded by the Drosophila melanogaster BAC clone BACR48M05 (AC005719) as prediction by the computer program Fgene.

```
Output file for REVERSE STRAND of FGene
F469BE1C
 length of sequence - 188357
 number of predicted exons - 21
 positions of predicted exons:
   4726 -    4757 w=   4.11  ORF:    4726 -    4755
   4816 -    4966 w=  20.57  ORF:    4817 -    4966
   5018 -    5318 w=  15.85  ORF:    5018 -    5317
   8693 -    8727 w=  14.75  ORF:    8695 -    8727
  38041 -   38265 w=   8.43  ORF:   38041 -   38265
  62411 -   62522 w=  10.60  ORF:   62411 -   62521
  74061 -   74692 w=  19.39  ORF:   74063 -   74692
 103484 -  103654 w=  24.14  ORF:  103484 -  103654
 132758 -  133134 w=  17.28  ORF:  132758 -  133132
 153576 -  153706 w=  18.42  ORF:  153577 -  153705
 154573 -  154681 w=  20.72  ORF:  154575 -  154679
 154753 -  156246 w=  23.66  ORF:  154754 -  156244
 160324 -  160375 w=   6.48  ORF:  160325 -  160375
 161337 -  161421 w=   6.82  ORF:  161337 -  161420
 171340 -  171756 w=  10.27  ORF:  171342 -  171755
 171821 -  171965 w=  18.76  ORF:  171823 -  171963
 172024 -  172326 w=  15.53  ORF:  172025 -  172324
 174437 -  174810 w=   9.70  ORF:  174438 -  174809
 175017 -  175168 w=  16.41  ORF:  175019 -  175168
 179216 -  179267 w=   6.89  ORF:  179216 -  179266
 187662 -  187678 w=   5.32  ORF:  187664 -  187678

Length of Coding region-   5367bp
Amino acid sequence -   1788aa

MDSGICYIKPEYLVTEADGGSAAANTENSDTNKRKREDGGEVEAGEKKKWDKKERKRGQN
KNRPVFKDERYSHLCHSLIDGTGGEPCSLANCRYVHDLDAYLAAKGEDLGPECYVYTTKG
YCARGVSCRFAKAHTDEQGRNLKREDYDENAPPTTCNGVSSAASSTLHNASMQMNPLTNM
KNVLKLSEHELQHGGKKSWHDMYKDSAWIFVAGFPYTLTEGDLVCVFSQYGEVVNINLIR
DSKTGKSKHSPLYRGEILFRIPELSQIPDPLCFLCNSIKLNSEVLNPANFPMDIGIPNPY
TNEQLVNAKLEQQNLEKLFNELENTASMSNSQESKDTETTSTALVESSTSTNSASSAGSC
SLANPAQQSMKKKLTFLNLSPFRSGKKSIDKNTSEQQRAISELVSTDHMLHLQQLLQQQR
KDQRSHTVPTESNYVLFNPGPVPSRHVQYKIRKPRPLSTHSDADSGFLSPCSPEEMRANP
AILVLQQCDSVQGYMEIYTDWANYYLERAKSKRKVTDLSADCRDGLLLAEVIEAVTSFKV
PDLVKKPKNQQQMFDNVNSCLHVLRSQSVGGLENITTNDICAGRLKAVLALFFALSRFKQ
QAKQTKSIGVGCGGGVGGSSSTLTGSGSVLGIGIGGLRTPGSSLNQDKNQQEQQQQQQQQ
QTPQQLAQSLENGNEMVNRQIAPAYAKVNGGTAIPLPATVMVQRRCPPDKVRPLPPTPNH
TPSIPGLGKSGSDFNTSRPNSPPTSNHTIQSLKSGNNNSLRPPSIKSGIPSPSSPQTAPQ
KHSMLDKLKLFNKEKQQNAVNAASVASKTQIQSKRTSSSSGFSSARSERSDSSLSLNDGH
GSQLKPPSISVSSQKPQPKTKQSKLLAAQQKKEQANKATKLDKKEKSPARSLNKEESGNE
SRSSTMGRTGKSSLVRAVGGVEKNTPKTSSKSSLHSKSDSKSSLKAPQLLQSPSSGGLPK
PIAAIKGTSKLPSLGGGAGHLPAAESQQNQQLLKRETSDISSNISQPPPAEPPISTHAHI
HQNQTPPPPYYANSQPTSHISSHGFLSEPSTPQHSSGIYGSSRLPPPKSALSAPRKLEYN
AGPHILSSPTHHQRQGLPRPLVNSAPNTPTASPNKFHTIPSKIVGTIYESKEEQLLPAPP
PASGGSSILPMRPLLRGYNSHVTLPTRGARGGHHPHQSYLDFCESDIGQGYCSDGDALRV
GSSPGGSRFHDIDNGYLSEGSSGLNGPSSSAGGISPGKHFLSMMRARTQLPTTIEERQLI
YGASVPILTLLPDRKIYQNNVRQIKVDKLAAMAERWNMELGNGGAKMDGSPHHRPGSRNG
RDNWSKMPEPLNGQKVEKSDKSSPSRRSMGGGGSGSSSKQGSPSSSSRTKGVPPSFGYVK
RANGSIASTAEQQNIAMMMAAGGAGANGLPCGRTAHVSAVPRTASGRKVAGGTQTLPNDM
NKLPPNTQHRSFSLTGPTATQLSQSIRERLATGSHSLPKPGSDLHVFQHRISNRGGTRHD
```

Figure 11b (CONTINUED)

GSLSDTQTYAEVKPEYSSYAMWLKHSNTAGSRLSDGESVEQLQIGSPALTRHGHKMIHNR
SGGPGQMAGQMSGNESPYVQSPRMNRSNSIRSTKSEKMYPSMMSRAGEVEIEPYYCLPVG
TNGVLTAQMAAAMAAQSQAAQGNPGVGVNVGGVAWSQPTSPTPLTRGPFNTAAGASVLSP
THGTTSAAGLVGPGGGAGGGAMVGHRLTYPKKNDEVHGSAASLLSGGSSLYGNAEERQAH
EIRRLKRELQDARDQVLSLSSQLSTNVSKKCPVVVFQMYTLRMARSRR*

Figure 11c: Illustration of a 'BLAST 2 sequences' search result with Hs-unc-53/3 as query and the Fgene predicted UNC53 homology ORF of Drosophila melanogaster BAC clone BACR48M05 as subject

Query: Hs-unc-53/3 (direct) 2120aa Length 2119 from:1 to = 2119
Subject: drosUNC53 (Fgene-prediction) Length 1788 from:1 to = 1788

Score = 106 bits (261), Expect = 2e-21
Identities = 190/840 (22%), Positives = 294/840 (34%), Gaps = 185/840 (22%)

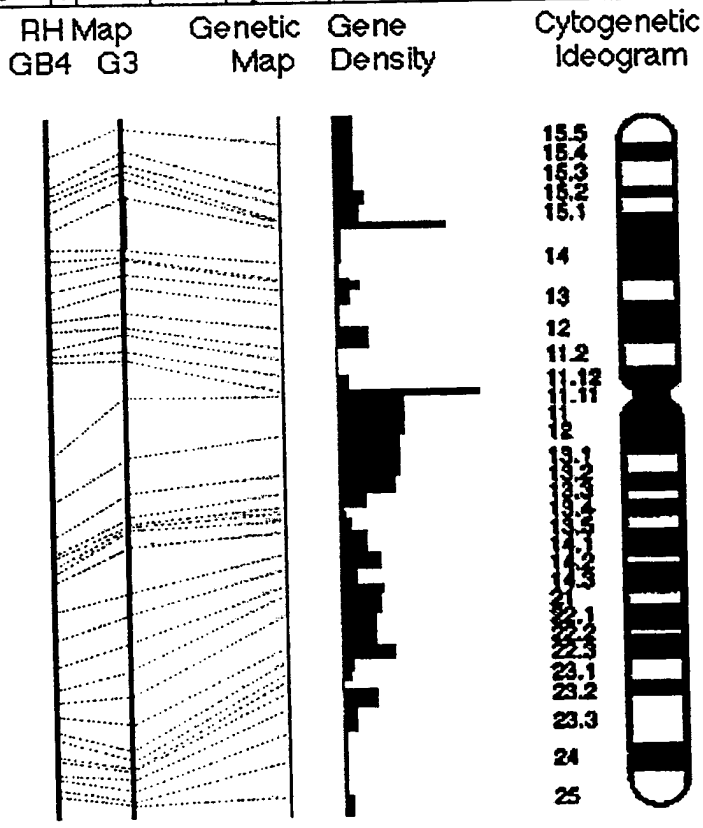
```
Query: 1    IYTDWANHYLAKSGHKRLIKDLQQDIADGVLLAEIIQIIANEKVEDINGCPRSQSQMIEN 60
             IYTDWAN+YL ++  KR + DL  D  DG+LLAE+I+ + + KV D+   P++Q QM +N
Sbjct: 497   IYTDWANYYLERAKSKRKVTDLSADCRDGLLLAEVIEAVTSFKVPDLVKKPKNQQQMFDN 556

Query: 61   VDVCLSFLAARGV-NVQGLSAEEIRNGNLKAILGLFFSLSRYK----------------- 102
             V+ CL  L ++ V  ++ ++  +I  G LKA+L LFF+LSR+K
Sbjct: 557   VNSCLHVLRSQSVGGLENITTNDICAGRLKAVLALFFALSRFKQQAKQTKSIGVGCGGGV 616

Query: 103  XXXXXXXXXXXSLVEL---QQRVTHASPPSEASQAKTQQDMQSSLAARYATQSNHSG--- 156
                      S++ +     R   +S   + +Q + QQ  Q      +   QS  +G
Sbjct: 617   GGSSSTLTGSGSVLGIGIGGLRTPGSSLNQDKNQQEQQQQQQQQQTPQQLAQSLENGNEM 676

Query: 157  ----IATSQKK---PTRLPGPSRV------------PAAGSSSKVQGASNLNRRSQSFNS 197
                 IA +  K    T +P P+ V                P   + +       L +   FN+
Sbjct: 677   VNRQIAPAYAKVNGGTAIPLPATVMVQRRCPPDKVRPLPPTPNHTPSIPGLGKSGSDFNT 736

Query: 198  IDKNKPPNYANGNEKDSSKGPQS-SSGVNGNVQPPSTAGQXXXXXXXXXXXXXKPWRSKSM 256
               N PP         S+   QS  SG N +++PPS
Sbjct: 737   SRPNSPPT--------SNHTIQSLKSGNNNSLRPPSIKSGI------------------ 769

Query: 257  NVKHSATSTMLTVKQXXXXXXXXXXXDRLKPPVSEGVKTAPSGQKSMLEKFKLVNARTAL 316
                                          P     +TAP  + SML+K KL N
Sbjct: 770   ------------------------------PSPSSPQTAPQ-KHSMLDKLKLFNKEKQQ 797

Query: 317  RXXXXXXXXXXXXXXXXXAFSESGEMEGFXXXXXXXXXXXXXPKVSPKLAPPKAGSKNLS 376
                              S SG                          +L PP   S ++S
Sbjct: 798   NAVNAASVASKTQIQSKRTSSSSGFSS--ARSERSDSSLSLNDGHGSQLKPP---SISVS 852

Query: 377  NKKSLLQPXXXXXXNRDKNKVCTEKPVKEEKDQVTEMAPKKTSKIASLIPKGSKTTAAKK 436
             ++K   QP          ++K+   +  KE+ ++ T++  K+ S   SL  + S   +  +
Sbjct: 853   SQKP--QP------KTKQSKLLAAQQKKEQANKATKLDKKEKSPARSLNKEESGNES--R 902

Query: 437  ESLXXXXXXXXXXXXXXXTVKQTISPGSTASKESEKFRTTKGSPSQSLSKPITMEKASAS 496
              S                    K T    S +S  S       K      SL  P  ++ ++S+
Sbjct: 903   SSTMGRTGKSSLVRAVGGVEKNTPKTSSKSSLHS------KSDSKSSLKAPQLLQSPSSG 956

Query: 497  SCPAPLEGREAGQASPSGSCTMTVAQSSGQSTGNGAVQLP-----QQQQHSHPNTATVA- 550
               P P+   +     P                S G GA  LP      Q QQ      T+ ++
Sbjct: 957   GLPKPIAAIKGTSKLP--------------SLGGGAGHLPAAESQQNQQLLKRETSDISS 1002

Query: 551  ----------PFIYRAHSENEGTALPSADSCTSPTKMDLSYSKTAKQCLEEISGEGPETR 600
                      P    AH      T  P   + + PT    S+   ++    + S    +
Sbjct: 1003  NISQPPPAEPPISTHAHIHQNQTPPPPYYANSQPTSHISSHGFLSEPSTPQHSSGIYGSS 1062

Query: 601  RMRTVKNIADLRQNLEETMSSLRGTQISHSTLETTFDSTVTTEVNGRTI-PN-LTSRPTP 658
             R+   K+     + LE          +  +H   +      V +  N  T  PN   + P+
Sbjct: 1063  RLPPPKSALSAPRKLEYNAGPHILSSPTHHQRQGLPRPLVNSAPNTPTASPNKFHTIPSK 1122

Query: 659  MTWRLGQACPRLQAGDAPSLGAGYPRSGTSRFIHTDPSRFMY----TTPLRRAAVSRLGN 714
             +   +       ++ +   L A  P SG S  +   P    Y    T P R A      +
Sbjct: 1123  IVGTI------YESKEEQLLPAPPPASGGSSILPMRPLLRGYNSHVTLPTRGARGGHHPH 1176

Query: 715  MSQIDMSEKASSDLDMSSEVDVG-GYMSDGDIL--GKS---LRTDDINSGYMTDG--GLN 766
              S +D  E             D+G GY SDGD L  G S     R  DI++GY+++G  GLN
Sbjct: 1177  QSYLDFCES-----------DIGQGYCSDGDALRVGSSPGGSRFHDIDNGYLSEGSSGLN 1225
```

Figure 12: Illustration of an EST encoding a part of the Zebrafish-UNC-53/2 cDNA.

```
Query= hh2UNC53          (2340 letters)

Sbjct= emb|AI658309|AI658309 fc21d06.y1 Zebrafish WashU MPIMG EST Danio
rerio cDNA 5' similar to TR:Q20427 Q20427 F45E10.1 mRNA sequence. Length = 445

 Score =  277 bits (702), Expect = 4e-73
 Identities = 124/147 (84%), Positives = 136/147 (92%)
 Frame = +3

Query: 2121 LHHNFRWVLCANHTEPVKGFLGRFLRRKLMETEISGRVRNMELVKIIDWIPKVWHHLNRF 2180
          LHHNFRW+LCANHTEPVKGFLGRFLRRKL+ETEI+ RVRN ELVKII+WIP VWHHLNRF
Sbjct: 3    LHHNFRWILCANHTEPVKGFLGRFLRRKLLETEINSRVRNGELVKIIEWIPSVWHHLNRF 182

Query: 2181 LEAHSSSDVTIGPRLFLSCPIDVDGSRVWFTDLWNYSIIPYLLEAVREGLQLYGRRAPWE 2240
          LE HSSSDVTIGPRLFLSCP+DV+GSRVWFTDLWNYSIIPY+LEAVREGLQ+YGR+A WE
Sbjct: 183  LETHSSSDVTIGPRLFLSCPMDVEGSRVWFTDLWNYSIIPYMLEAVREGLQMYGRKASWE 362

Query: 2241 DPAKWVMDTYPWAASPQQHEWPPLLQL 2267
          DPAKWVM++    ASPQQHEW  LL+L
Sbjct: 363  DPAKWVMESLLCVASPQQHEWHSLLRL 443
```

Figure 13. Genemap98 results for Hs-Unc53/2

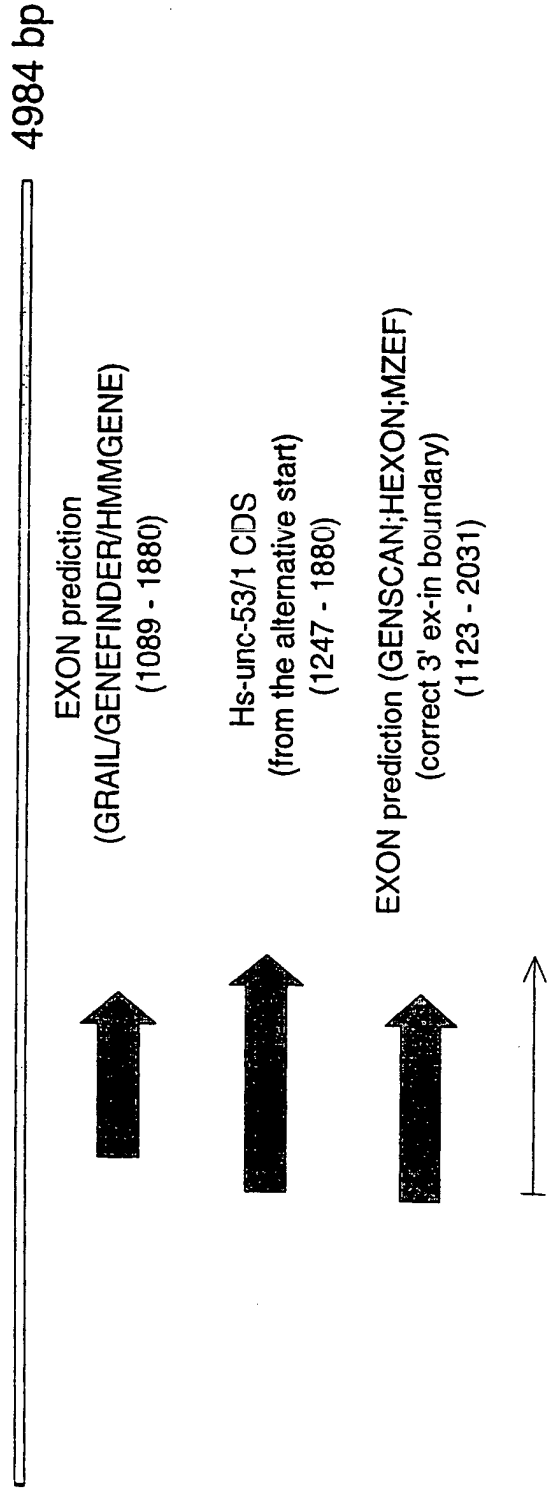
UniGene	Hs.13830						
RH Mapping Results							
SHGC-33456	G3 Map:	Chr.11					
	Reference interval:	D11S921-D11S1359 (24.9-32.5 cM)					
	Physical position:	911 cR10000 (F)					
	RH details:	RHdb RH32790					
	Typed by:	Stanford (see SHGC-33456)					
Electronic PCR Results							
ESTs (from GenBank EST division)							
AA115015	zl04d10.s1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone 491347 3'						
	STS	7	...	134	bp:	SHGC-33456	
AA918601	ol53e11.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:1527212 3'						
	STS	16	...	143	bp:	SHGC-33456	
AI248585	qh71f08.x1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA clone IMAGE:1850151 3', mRNA sequence [Homo sapiens]						
	STS	19	...	146	bp:	SHGC-33456	
T71262	yd35b09.s1 Homo sapiens cDNA clone 110201 3'.						
	STS	9	...	136	bp:	SHGC-33456	



The thick line on the G3 map indicates the position of SHGC-33456 See also: equivalent interval on GB4 map

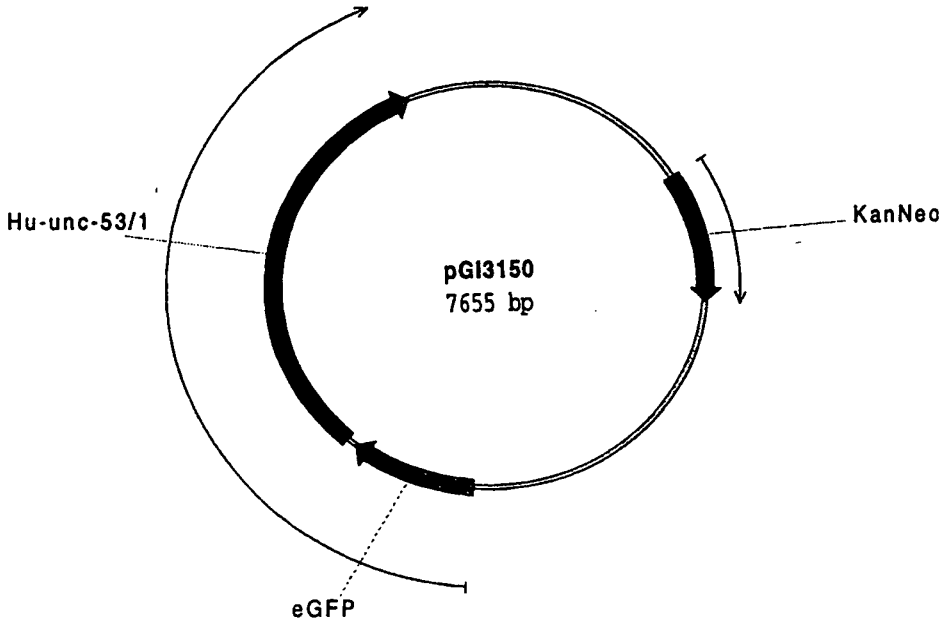
About This Interval	
Top of interval:	D11S921 (24.9 cM)
Bottom of interval:	D11S1359 (32.5 cM)
Genetic size of bin:	8 cM
Physical size of bin:	430 cR10000

Figure 14. Prediction of a 5' exon of Hs-unc-53/1*



(*) numbers refer to figure 1g.

Figure 15: Illustration of the nucleotide sequence of pGI3150 and amino acid sequence of the eGFP fusion with a C-terminal fragment of Hs-Unc-53/1.




```
ID   pGI3150                          circular DNA; 7655 BP
DE   from coiled coil I till end
FT   CDS             1225..2019
FT                   /vntifkey="4"
FT                   /label=KanNeo
FT   CDS             3942..4658
FT                   /vntifkey="4"
FT                   /label=eGFP
FT   CDS             4719..7214
FT                   /vntifkey="4"
FT                   /label=Hu-unc-53/1
SQ   SEQUENCE   7655 BP;
```

```
CTAGATAACT GATCATAATC AGCCATACCA CATTTGTAGA GGTTTTACTT GCTTTAAAAA      60
ACCTCCCACA CCTCCCCCTG AACCTGAAAC ATAAAATGAA TGCAATTGTT GTTGTTAACT     120
TGTTTATTGC AGCTTATAAT GGTTACAAAT AAAGCAATAG CATCACAAAT TTCACAAATA     180
AAGCATTTTT TTCACTGCAT TCTAGTTGTG GTTTGTCCAA ACTCATCAAT GTATCTTAAC     240
GCGTAAATTG TAAGCGTTAA TATTTTGTTA AAATTCGCGT TAAATTTTTG TTAAATCAGC     300
TCATTTTTTA ACCAATAGGC CGAAATCGGC AAAATCCCTT ATAAATCAAA AGAATAGACC     360
GAGATAGGGT TGAGTGTTGT TCCAGTTTGG AACAAGAGTC CACTATTAAA GAACGTGGAC     420
TCCAACGTCA AAGGGCGAAA AACCGTCTAT CAGGGCGATG GCCCACTACG TGAACCATCA     480
CCCTAATCAA GTTTTTTGGG GTCGAGGTGC CGTAAAGCAC TAAATCGGAA CCCTAAAGGG     540
AGCCCCCGAT TTAGAGCTTG ACGGGGAAAG CCGGCGAACG TGGCGAGAAA GGAAGGGAAG     600
AAAGCGAAAG GAGCGGGCGC TAGGGCGCTG GCAAGTGTAG CGGTCACGCT GCGCGTAACC     660
ACCACACCCG CCGCGCTTAA TGCGCCGCTA CAGGGCGCGT CAGGTGGCAC TTTTCGGGGA     720
AATGTGCGCG GAACCCCTAT TTGTTTATTT TTCTAAATAC ATTCAAATAT GTATCCGCTC     780
ATGAGACAAT AACCCTGATA AATGCTTCAA TAATATTGAA AAAGGAAGAG TCCTGAGGCG     840
GAAAGAACCA GCTGTGGAAT GTGTGTCAGT TAGGGTGTGG AAAGTCCCCA GGCTCCCCAG     900
CAGGCAGAAG TATGCAAAGC ATGCATCTCA ATTAGTCAGC AACCAGGTGT GGAAAGTCCC     960
CAGGCTCCCC AGCAGGCAGA AGTATGCAAA GCATGCATCT CAATTAGTCA GCAACCATAG    1020
TCCCGCCCCT AACTCCGCCC ATCCCGCCCC TAACTCCGCC CAGTTCCGCC CATTCTCCGC    1080
CCCATGGCTG ACTAATTTTT TTTATTTATG CAGAGGCCGA GGCCGCCTCG GCCTCTGAGC    1140
TATTCCAGAA GTAGTGAGGA GGCTTTTTTG GAGGCCTAGG CTTTTGCAAA GATCGATCAA    1200
GAGACAGGAT GAGGATCGTT TCGCATGATT GAACAAGATG GATTGCACGC AGGTTCTCCG    1260
GCCGCTTGGG TGGAGAGGCT ATTCGGCTAT GACTGGGCAC AACAGACAAT CGGCTGCTCT    1320
GATGCCGCCG TGTTCCGGCT GTCAGCGCAG GGGCGCCCGG TTCTTTTTGT CAAGACCGAC    1380
CTGTCCGGTG CCCTGAATGA ACTGCAAGAC GAGGCAGCGC GGCTATCGTG GCTGGCCACG    1440
ACGGGCGTTC CTTGCGCAGC TGTGCTCGAC GTTGTCACTG AAGCGGGAAG GGACTGGCTG    1500
```

Figure 15 (CONTINUED 1)

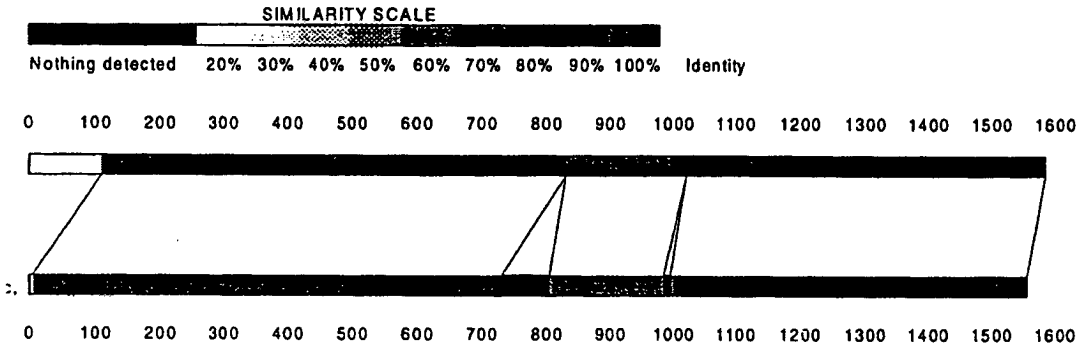
```
CTATTGGGCG AAGTGCCGGG GCAGGATCTC CTGTCATCTC ACCTTGCTCC TGCCGAGAAA 1560
GTATCCATCA TGGCTGATGC AATGCGGCGG CTGCATACGC TTGATCCGGC TACCTGCCCA 1620
TTCGACCACC AAGCGAAACA TCGCATCGAG CGAGCACGTA CTCGGATGGA AGCCGGTCTT 1680
GTCGATCAGG ATGATCTGGA CGAAGAGCAT CAGGGGCTCG CGCCAGCCGA ACTGTTCGCC 1740
AGGCTCAAGG CGAGCATGCC CGACGGCGAG GATCTCGTCG TGACCCATGG CGATGCCTGC 1800
TTGCCGAATA TCATGGTGGA AAATGGCCGC TTTTCTGGAT TCATCGACTG TGGCCGGCTG 1860
GGTGTGGCGG ACCGCTATCA GGACATAGCG TTGGCTACCC GTGATATTGC TGAAGAGCTT 1920
GGCGGCGAAT GGGCTGACCG CTTCCTCGTG CTTTACGGTA TCGCCGCTCC CGATTCGCAG 1980
CGCATCGCCT TCTATCGCCT TCTTGACGAG TTCTTCTGAG CGGGACTCTG GGGTTCGAAA 2040
TGACCGACCA AGCGACGCCC AACCTGCCAT CACGAGATTT CGATTCCACC GCCGCCTTCT 2100
ATGAAAGGTT GGGCTTCGGA ATCGTTTTCC GGGACGCCGG CTGGATGATC CTCCAGCGCG 2160
GGGATCTCAT GCTGGAGTTC TTCGCCCACC CTAGGGGGAG GCTAACTGAA ACACGGAAGG 2220
AGACAATACC GGAAGGAACC CGCGCTATGA CGGCAATAAA AAGACAGAAT AAAACGCACG 2280
GTGTTGGGTC GTTTGTTCAT AAACGCGGGG TTCGGTCCCA GGGCTGGCAC TCTGTCGATA 2340
CCCCACCGAG ACCCCATTGG GGCCAATACG CCCGCGTTTC TTCCTTTTCC CCACCCCACC 2400
CCCCAAGTTC GGGTGAAGGC CCAGGGCTCG CAGCCAACGT CGGGGCGGCA GGCCCTGCCA 2460
TAGCCTCAGG TTACTCATAT ATACTTTAGA TTGATTTAAA ACTTCATTTT TAATTTAAAA 2520
GGATCTAGGT GAAGATCCTT TTTGATAATC TCATGACCAA AATCCCTTAA CGTGAGTTTT 2580
CGTTCCACTG AGCGTCAGAC CCCGTAGAAA AGATCAAAGG ATCTTCTTGA GATCCTTTTT 2640
TTCTGCGCGT AATCTGCTGC TTGCAAACAA AAAAACCACC GCTACCAGCG GTGGTTTGTT 2700
TGCCGGATCA AGAGCTACCA ACTCTTTTTC CGAAGGTAAC TGGCTTCAGC AGAGCGCAGA 2760
TACCAAATAC TGTCCTTCTA GTGTAGCCGT AGTTAGGCCA CCACTTCAAG AACTCTGTAG 2820
CACCGCCTAC ATACCTCGCT CTGCTAATCC TGTTACCAGT GGCTGCTGCC AGTGGCGATA 2880
AGTCGTGTCT TACCGGGTTG GACTCAAGAC GATAGTTACC GGATAAGGCG CAGCGGTCGG 2940
GCTGAACGGG GGGTTCGTGC ACACAGCCCA GCTTGGAGCG AACGACCTAC ACCGAACTGA 3000
GATACCTACA GCGTGAGCTA TGAGAAAGCG CCACGCTTCC CGAAGGGAGA AAGGCGGACA 3060
GGTATCCGGT AAGCGGCAGG GTCGGAACAG GAGAGCGCAC GAGGGAGCTT CCAGGGGGAA 3120
ACGCCTGGTA TCTTTATAGT CCTGTCGGGT TTCGCCACCT CTGACTTGAG CGTCGATTTT 3180
TGTGATGCTC GTCAGGGGGG CGGAGCCTAT GGAAAAACGC CAGCAACGCG GCCTTTTTAC 3240
GGTTCCTGGC CTTTTGCTGG CCTTTTGCTC ACATGTTCTT TCCTGCGTTA TCCCCTGATT 3300
CTGTGGATAA CCGTATTACC GCCATGCATT AGTTATTAAT AGTAATCAAT TACGGGGTCA 3360
TTAGTTCATA GCCCATATAT GGAGTTCCGC GTTACATAAC TTACGGTAAA TGGCCCGCCT 3420
GGCTGACCGC CCAACGACCC CCGCCCATTG ACGTCAATAA TGACGTATGT TCCCATAGTA 3480
ACGCCAATAG GGACTTTCCA TTGACGTCAA TGGGTGGAGT ATTTACGGTA AACTGCCCAC 3540
TTGGCAGTAC ATCAAGTGTA TCATATGCCA AGTACGCCCC CTATTGACGT CAATGACGGT 3600
AAATGGCCCG CCTGGCATTA TGCCCAGTAC ATGACCTTAT GGGACTTTCC TACTTGGCAG 3660
TACATCTACG TATTAGTCAT CGCTATTACC ATGGTGATGC GGTTTTGGCA GTACATCAAT 3720
GGGCGTGGAT AGCGGTTTGA CTCACGGGGA TTTCCAAGTC TCCACCCCAT TGACGTCAAT 3780
GGGAGTTTGT TTTGGCACCA AAATCAACGG GACTTTCCAA AATGTCGTAA CAACTCCGCC 3840
CCATTGACGC AAATGGGCGG TAGGCGTGTA CGGTGGGAGG TCTATATAAG CAGAGCTGGT 3900
TTAGTGAACC GTCAGATCCG CTAGCGCTAC CGGTCGCCAC CATGGTGAGC AAGGGCGAGG 3960
AGCTGTTCAC CGGGGTGGTG CCCATCCTGG TCGAGCTGGA CGGCGACGTA AACGGCCACA 4020
AGTTCAGCGT GTCCGGCGAG GGCGAGGGCG ATGCCACCTA CGGCAAGCTG ACCCTGAAGT 4080
TCATCTGCAC CACCGGCAAG CTGCCCGTGC CCTGGCCCAC CCTCGTGACC ACCCTGACCT 4140
ACGGCGTGCA GTGCTTCAGC CGCTACCCCG ACCACATGAA GCAGCACGAC TTCTTCAAGT 4200
CCGCCATGCC CGAAGGCTAC GTCCAGGAGC GCACCATCTT CTTCAAGGAC GACGGCAACT 4260
ACAAGACCCG CGCCGAGGTG AAGTTCGAGG GCGACACCCT GGTGAACCGC ATCGAGCTGA 4320
AGGGCATCGA CTTCAAGGAG GACGGCAACA TCCTGGGGCA CAAGCTGGAG TACAACTACA 4380
ACAGCCACAA CGTCTATATC ATGGCCGACA AGCAGAAGAA CGGCATCAAG GTGAACTTCA 4440
AGATCCGCCA CAACATCGAG GACGGCAGCG TGCAGCTCGC CGACCACTAC CAGCAGAACA 4500
CCCCCATCGG CGACGGCCCC GTGCTGCTGC CCGACAACCA CTACCTGAGC ACCCAGTCCG 4560
CCCTGAGCAA AGACCCCAAC GAGAAGCGCG ATCACATGGT CCTGCTGGAG TTCGTGACCG 4620
CCGCCGGGAT CACTCTCGGC ATGGACGAGC TGTACAAGTC CGGACTCAGA TCTCGAGCTC 4680
AAGCTTCGAA TTCTGCAGTC GACGGTACCG CGGGCCCGGG ATCCTTCCGA GACCCCACGG 4740
ACGATGTTCA CGGCTCAGTG CTGTCCCTGG CCTCCAGTGC CTCCTCCACC TACTCCTCAG 4800
CTGAGGAGAG GATGCAATCT GAGCAAATCC GGAAGCTTCG TAGGGAACTG GAATCATCCC 4860
AGGAAAAAGT GGCCACCTTG ACGTCTCAGC TTTCTGCCAA TGCTAATCTG GTGGCTGCTT 4920
TTGAGCAGAG CCTGGTGAAT ATGACATCCC GCCTGCGACA CCTGGCAGAG ACGGCCGAGG 4980
AGAAGGACAC TGAGCTGCTG GATTTGCGAG AAACCATAGA CTTTCTGAAG AAAAAGAACT 5040
CTGAGGCCCA GGCAGTCATT CAGGGAGCCC TTAATGCCTC AGAAACCACA CCCAAAGAAC 5100
TTCGGATCAA GAGACAAAAC TCCTCAGATA GCATCTCAAG CCTCAACAGC ATCACTAGCC 5160
ATTCCAGCAT CGGCAGCAGC AAGGATGCTG ATGCGAAAAA GAAGAAAAAA AAGAGTTGGG 5220
TCTATGAGCT TCGAAGTTCC TTCAACAAAG CGTTCAGTAT AAAAAAGGGG CCCAAGTCAG 5280
CTTCCTCATA CTCGGATATA GAGGAGATTG CTACACCCGA CTCTTCAGCC CCCTCATCCC 5340
CCAAACTACA GCATGGTTCC ACAGAGACTG CTTCACCCTC CATCAAGTCC TCCACCTTGT 5400
CCTCCGTGGG CACTGATGTC ACCGAGGGCC CTGCTCACCC AGCCCCCCAC ACTAGGCTGT 5460
TCCATGCAAA TGAGGAGGAG GAGCCAGAGA AGAAGGAGGT ATCGGAGCTG CGCTCTGAGC 5520
TATGGGAGAA GGAAATGAAG CTTACAGACA TCCGCTTGGA GGCCCTCAAC TCTGCCCACC 5580
AACTGGATCA GCTTCGGGAG ACCATGCACA ACATGCAGTT GGAGGTGGAC CTGCTGAAAG 5640
CAGAGAATGA CCGACTGAAG GTAGCCCCAG GCCCCTCATC AGGCTCCACT CCAGGGCAGG 5700
TCCCTGGATC ATCTGCATTA TCTTCCCCAC GCCGCTCCCT AGGCCTGGCA CTCACCCATT 5760
CCTTCGGCCC CAGTCTTGCA GACACAGACC TGTCACCCAT GGATGGCATC AGTACTTGTG 5820
```

Figure 15 (CONTINUED 2)

```
GTCCAAAGGA GGAAGTGACC CTCCGGGTGG TGGTGAGGAT GCCCCCGCAG CACATCATCA      5880
AAGGGGACTT GAAGCAGCAG GAATTCTTCC TGGGCTGTAG CAAGGTCAGT GGAAAAGTTG      5940
ACTGGAAGAT GCTGGATGAA GCTGTTTTCC AAGTGTTCAA GGACTATATT TCTAAAATGG      6000
ACCCAGCCTC TACCCTGGGA CTAAGCACTG AGTCCATCCA TGGCTACAGC ATCAGCCACG      6060
TGAAACGAGT GTTGGATGCA GAGCCCCCCG AGATGCCTCC TTGCCGTCGA GGTGTCAATA      6120
ACATATCAGT CTCCCTCAAA GGTCTGAAGG AGAAATGCGT CGACAGCCTG GTGTTCGAGA      6180
CGCTGATCCC CAAGCCGATG ATGCAGCACT ACATAAGCCT CCTGCTGAAG CACCGGCGCC      6240
TCGTCCTCTC GGGCCCCAGC GGCACGGGCA AGACCTACCT GACCAATCGC TTGGCCGAGT      6300
ACCTGGTGGA GCGCTCTGGC CGTGAGGTCA CAGAGGGCAT CGTCAGCACC TTCAACATGC      6360
ACCAGCAGTC TTGCAAGGAT CTGCAACTGT ATCTTTCCAA CCTAGCCAAC CAGATAGACC      6420
GGGAAACAGG AATTGGGGAT GTGCCCCTGG TGATTCTATT GGATGACCTG AGTGAAGCAG      6480
GCTCCATCAG TGAGTTGGTC AATGGGGCCC TCACCTGCAA GTATCATAAA TGTCCCTATA      6540
TTATAGGTAC CACCAATCAG CCTGTAAAAA TGACACCCAA CCATGGCTTG CACTTGAGCT      6600
TCAGGATGTT GACCTTCTCC AACAACGTGG AGCCAGCCAA TGGCTTCCTG GTTCGTTACC      6660
TGAGGAGGAA GCTGGTAGAG TCAGACAGCG ACATCAATGC CAACAAGGAA GAGCTGCTTC      6720
GGGTGCTCGA CTGGGTACCC AAGCTGTGGT ATCATCTCCA CACCTTCCTT GAGAAGCACA      6780
GCACCTCAGA CTTCCTCATC GGCCCTTGCT TCTTTCTGTC GTGTCCCATT GGCATTGAGG      6840
ACTTCCGGAC CTGGTTCATT GACCTGTGGA ACAACTCTAT CATTCCCTAT CTACAGGAAG      6900
GAGCCAAGGA TGGGATAAAG GTCCATGGAC AGAAAGCTGC TTGGGAGGAC CCAGTGGAAT      6960
GGGTCCGGGA CACACTTCCC TGGCCATCAG CCCAACAAGA CCAATCAAAG CTGTACCACC      7020
TGCCCCCACC CACCGTGGGC CCTCACAGCA TTGCCTCACC TCCCGAGGAT AGGACAGTCA      7080
AAGACAGCAC CCCAAGTTCT CTGGACTCAG ATCCTCTGAT GGCCATGCTG CTGAAACTTC      7140
AAGAAGCTGC CAACTACATT GAGTCTCCAG ATCGAGAAAC CATCCTGGAC CCCAACCTTC      7200
AGGCAACACT TTAAGGGTTC GGCAATCACT GTCACCCCCG GACAGCAGAA CGCTGGCATC      7260
AGCTATCTTA GCTCCTCCTC TCCCCTCTCC TCTTTCAGAG CACTGGCTCT CCAGCCCCAG      7320
GAGGAGAACA GGAGGGAGGA GGAGATGAAA GAGGAGGGAC AGGTTCTTGG TGCTGTACCT      7380
TTGAGAACTT CCTAGGAAGG AATGGTGGGG TGGCGTTTGG GAACTTGTGC CCCCTAAACA      7440
CATTTACTGG CCTCCTCTAA TGACTTTGGG GAAAAGATGA TTCTGGGTCT TTCCCTTGAC      7500
TTCTTGTTTC AATTACAAAC TCCTGGGCTT TCTGGGGAGG GGTTCAGAAA ACATCAAAAC      7560
ACTGCAGCAG TTCCCCGGAA TTCAGCTTGG ACTTAACCAG GCTGAACTTG CTCAAAAGAA      7620
GCCGAATTCC AGCACACTGG CGGCCGTTAC TAGTT                                  7655
//
```

```
MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQC
FSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKL
EYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPN
EKRDHMVLLEFVTAAGITLGMDELYKSGLRSRAQASNSAVDGTAGPGSFRDPTDDVHGSVLSLASSASSTY
SSAEERMQSEQIRKLRRELESSQEKVATLTSQLSANANLVAAFEQSLVNMTSRLRHLAETAEEKDTELLDL
RETIDFLKKKNSEAQAVIQGALNASETTPKELRIKRQNSSDSISSLNSITSHSSIGSSKDADAKKKKKKSW
VYELRSSFNKAFSIKKGPKSASSYSDIEEIATPDSSAPSSPKLQHGSTETASPSIKSSTLSSVGTDVTEGP
AHPAPHTRLFHANEEEEPEKKEVSELRSELWEKEMKLTDIRLEALNSAHQLDQLRETMHNMQLEVDLLKAE
NDRLKVAPGPSSGSTPGQVPGSSALSSPRRSLGLALTHSFGPSLADTDLSPMDGISTCGPKEEVTLRVVVR
MPPQHIIKGDLKQQEFFLGCSKVSGKVDWKMLDEAVFQVFKDYISKMDPASTLGLSTESIHGYSISHVKRV
LDAEPPEMPPCRRGVNNISVSLKGLKEKCVDSLVFETLIPKPMMQHYISLLLKHRRLVLSGPSGTGKTYLT
NRLAEYLVERSGREVTEGIVSTFNMHQQSCKDLQLYLSNLANQIDRETGIGDVPLVILLDDLSEAGSISEL
VNGALTCKYHKCPYIIGTTNQPVKMTPNHGLHLSFRMLTFSNNVEPANGFLVRYLRRKLVESDSDINANKE
ELLRVLDWVPKLWYHLHTFLEKHSTSDFLIGPCFFLSCPIGIEDFRTWFIDLWNNSIIPYLQEGAKDGIKV
HGQKAAWEDPVEWVRDTLPWPSAQQDQSKLYHLPPPTVGPHSIASPPEDRTVKDSTPSSLDSDPLMAMLLK
LQEAANYIESPDRETILDPNLQATL
```

Figure 16: EST Clone yk480b6 contains a splice variant of Ce-UNC-53



Results of SIM with:

Sequence 1: Ce-unc-53, (1583 residues) Sequence 2: yk480b06rc, (1556 residues)

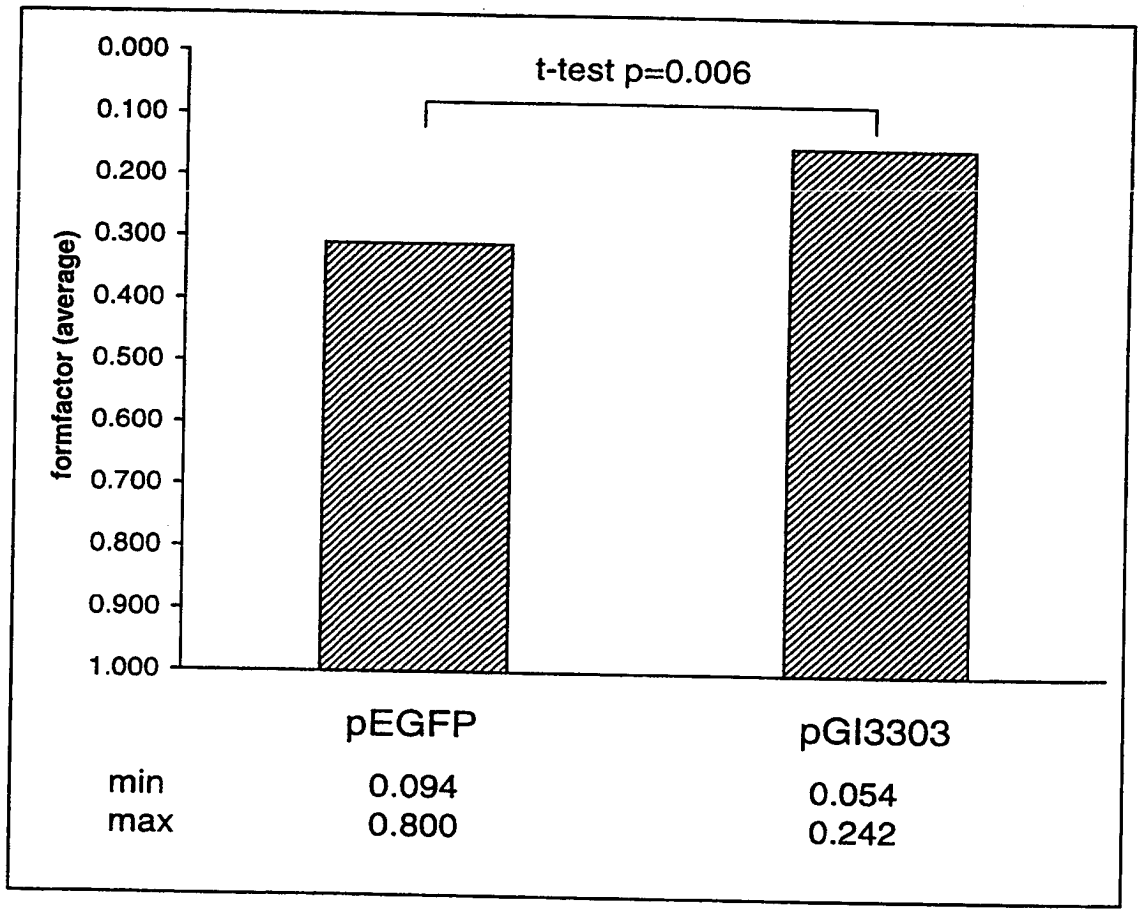
```
Ce-UNC-53     110 LSTYKQKLRQLKKDQKKLEQLPTSIMPPAVSKLPSPRVATSATASATNPNSNFPQMSTSR
yk480b06rc      5 LQEFGTRLRQLKKDQKKLEQLPTSIMPPAVSKLPSPRVATSATASATNPNSNFPQMSTSR
                  *          **************************************************

Ce-UNC-53     170 LQTPQSRISKIDSSKIGIKPKTSGLKPPSSSTTSSNNTNSFRPSSRSSGNNNVGSTISTS
yk480b06rc     65 LQTPQSRISKIDSSKIGIKPKTSGLKPPSSSTTSSNNTNSFRPSSRSSGNNNVGSTISTS
                  ************************************************************

Ce-UNC-53     230 AKSLESSSTYSSISNLNRPTSQLQKPSRPQTQLVRVATTTKIGSSKLAAPKAVSTPKLAS
yk480b06rc    125 AKSLESSSTYSSISNLNRPTSQLQKPSRPQTQLVRVATTTKIGSSKLAAPKAVSTPKLAS
                  ************************************************************

Ce-UNC-53     290 VKTIGAKQEPDNSGGGGGGMLKLKLFSSKNPSSSSNSPQPTRKAAAVPQQQTLSKIAAPV
yk480b06rc    185 VKTIGAKQEPDNSGGGGGGMLKLKLFSSKNPSSSSNSPQPTRKAAAVPQQQTLSKIAAPV
                  ************************************************************

Ce-UNC-53     350 KSGLKPPTSKLGSATSMSKLCTPKVSYRKTDAPIISQQDSKRCSKSSEEESGYAGFNSTS
yk480b06rc    245 KSGLKPPTSKLGSATSMSKLCTPKVSYRKTDAPIISQQDSKRCSKSSEEESGYAGFNSTS
                  ************************************************************

Ce-UNC-53     410 PTSSSTEGSLSMHSTSSKSSTSDEKSPSSDDLTLNASIVTAIRQPIAATPVSPNIIINKPV
yk480b06rc    305 PTSSSTEGSLSMHSTSSKSSTSDEKSPSSDDLTLNASIVTAIRQPIAATPVSPNIIINKPV
                  ************************************************************

Ce-UNC-53     470 EEKPTLAVKGVKSTAKKDPPPAVPPRDTQPTIGVVSPIMAHKKLTNDPVISEKPEPEKLQ
yk480b06rc    365 EEKPTLAVKGVKSTAKKDPPPAVPPRDTQPTIGVVSPIMAHKKLTNDPVISEKPEPEKLQ
                  ************************************************************

Ce-UNC-53     530 SMSIDTTDVPPLPPLKSVVPLKMTSIRQPPTYDVLLKQGKITSPVKSFGYEQSSASEDSI
yk480b06rc    425 SMSIDTTDVPPLPPLKSVVPLKMTSIRQPPTYDVLLKQGKITSPVKSFGYEQSSASEDSI
                  ************************************************************

Ce-UNC-53     590 VAHASAQVTPPTKTSGNHSLERRMGKNKTSESSGYTSDAGVAMCAKMREKLKEYDDMTRR
yk480b06rc    485 VAHASAQVTPPTKTSGNHSLERRMGKNKTSESSGYTSDAGVAMCAKMREKLKEYDDMTRR
                  ************************************************************


Ce-UNC-53     650 AQNGYPDNFEDSSSLSSGISDNNELDDISTDDLSGVDMATVASKHSDYSHFVRHPTSSSS
yk480b06rc    545 AQNGYPDNFEDSSSLSSGISDNNELDDISTDDLSGVDMATVASKHSDYSHFVRHPTSSSS
                  ************************************************************


Ce-UNC-53     710 KPRVPSRSSTSVDSRSRAEQENVYKLLSQCRTSQRGAAATSTFGQHSLRSPGYSSYSPHL
yk480b06rc    605 KPRVPSRSSTSVDSRSRAEQENVYKLLSQCRTSQRGAAATSTFGQHSLRSPGYSSYSPHL
                  ************************************************************

Ce-UNC-53     770 SVSADKDTMSMHSQTSRRPSSQKPSYSGQFHSLDRKCHLQEFTSTEHRMAALLSPRRVPN
yk480b06rc    665 SVSADKDTMSMHSQTSRRPSSQKPSYSGQFHSLDRKCHLQEFTSTEHRMAALLSPRRVPN
                  ************************************************************
```

Figure 16 (CONTINUED)

```
Ce-UNC-53     830 SMSKYDSS----------------------------------------------------------
yk480b06rc    725 SMSKYDSSAAALNASGMSRSMILLESLSPRPPRRHQSPADSCIITASPSAPRRSHSPRGP
                  ********

Ce-UNC-53     838 --------------------GSYSARSRGGSSTGIYGETFQLHRLSDEKSPAHSAKSEMGS
yk480b06rc    785 TARIPLSLASSPVHVNNNWGSYSARSRGGSSTGIYGETFQLHRLSDEKSPAHSAKSEMGS
                  ****************************************************

Ce-UNC-53     879 QLSLASTTAYGSLNEKYEHAIRDMARDLECYKNTVDSLTKKQENYGALFDLFEQKLRKLT
yk480b06rc    845 QLSLASTTAYGSLNEKYEHAIRDMARDLECYKNTVDSLTKKQENYGALFDLFEQKLRKLT
                  ************************************************************

Ce-UNC-53     939 QHIDRSNLKPEEAIRFRQDIAHLRDISNHLASNSAHANEGAGELLRQPSLESVASHRSSM
yk480b06rc    905 QHIDRSNLKPEEAIRFRQDIAHLRDISNHLASNSAHANEGAGELLRQPSLESVASHRSSM
                  ************************************************************

Ce-UNC-53     999 SSSSKSSKQEKISLSSFGKNKKSW-------IRSSLSKFTKKKNKNYDEAHMPSISGSQG
yk480b06rc    965 SSSSKSSKQEKISLSSFGKNKKSWALSVDSQIRSSLSKFTKKKNKNYDEAHMPSISGSQG
                  ************************ ******************************

Ce-UNC-53    1052 TLDNIDVIELKQELKERDSALYEVRLDNLDRAREVDVLRETVNKLKTENKQLKKEVDKLT
yk480b06rc   1025 TLDNIDVIELKQELKERDSALYEVRLDNLDRAREVDVLRETVNKLKTENKQLKKEVDKLT
                  ************************************************************

Ce-UNC-53    1112 NGPATRASSRASIPVIYDDEHVYDAACSSTSASQSSKRSSGCNSIKVTVNVDIAGEISSI
yk480b06rc   1085 NGPATRASSRASIPVIYDDEHVYDAACSSTSASQSSKRSSGCNSIKVTVNVDIAGEISSI
                  ************************************************************

Ce-UNC-53    1172 VNPDKEIIVGYLAMSTSQSCWKDIDVSILGLFEVYLSRIDVEHQLGIDARDSILGYQIGE
yk480b06rc   1145 VNPDKEIIVGYLAMETSQSCWKDIDVSILGLFEVYLSRIDVEHQLGIDARDSILGYQIGE
                  ************** ********************************************

Ce-UNC-53    1232 LRRVIGDSTTMITSHPTDILTSSTTIRMFMHGAAQSRVDSLVLDMLLPKQMILQLVKSIL
yk480b06rc   1205 LRRVIGDSTTMITSHPTDILTSSTTIRMFMHGAAQSRVDSLVLDMLLPKQMILQLVKSIL
                  ************************************************************

Ce-UNC-53    1292 TERRLVLAGATGIGKSKLAKTLAAYVSIRTNQSEDSIVNISIPENNKEELLQVERRLEKI
yk480b06rc   1265 TERRLVLAGATGIGKSKLAKTLAAYVSIRTNQSEDSIVNISIPENNKEELLQVERRLEKI
                  ************************************************************

Ce-UNC-53    1352 LRSKESCIVILDNIPKNRIAFVVSVFANVPLQNNEGPFVVCTVNRYQIPELQIHHNFKMS
yk480b06rc   1325 LRSKESCIVILDNIPKNRIAFVVSVFANVPLQNNEGPFVVCTVNRYQIPELQIHHNFKMS
                  ************************************************************

Ce-UNC-53    1412 VMSNRLEGFILRYLRRRAVEDEYRLTVQMPSELFKIIDFFPIALQAVNNFIEKTNSVDVT
yk480b06rc   1385 VMSNRLEGFILRYLRRRAVEDEYRLTVQMPSELFKIIDFFPIALQAVNNFIEKTNSVDVT
                  ************************************************************

Ce-UNC-53    1472 VGPRACLNCPLTVDGSREWFIRLWNENFIPYLERVARDGKKTFGRCTSFEDPTDIVSKKW
yk480b06rc   1445 VGPRACLNCPLTVDGSREWFIRLWNENFIPYLERVARDGKKTFGRCTSFEDPTDIVSEKW
                  ************************************************************

Ce-UNC-53    1532 PWFDGENPENVLKRLQLQDLVPSPANSSRQHFNPLESLIQLHATKHQTIDNI
yk480b06rc   1505 PWFDGENPENVLKRLQLQDLVPSPANSSRQHFNPLESLIQLHATKHQTIDNI
                  ****************************************************
```

Legend: the alternative splices and the mutation (S-P) are indicated in red and are boxed.

Figure 17.



INTERNATIONAL SEARCH REPORT

Interr[redacted]lication No
PCT/EP 99/03848

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/12 C07K14/47

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 96 38555 A (BOGAERT THIERRY ET AL.) 5 December 1996 (1996-12-05) page 1 -page 99; claims 1-88 ---	1-31, 49-54, 60-67, 72-80, 85,88, 91-95
A	HEKIMI S ET AL: "AXONAL GUIDANCE DEFECTS IN A CAENORHABDITID ELEGANS MUTANT REVEAL CELL-EXTRINSIC DETERMINANTS OF NEURONAL MORPHOLOGY" JOURNAL OF NEUROSCIENCE, vol. 13, no. 10, 1 October 1993 (1993-10-01), pages 4254-4271, XP000612286 ISSN: 0270-6474 the whole document --- -/--	1,21-26

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
5 November 1999	22/11/1999

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	De Kok, A

1

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	BAIROCH A.: "The PROSITE dictionary of sites and patterns in proteins, its current status" NUCLEIC ACIDS RESEARCH., vol. 21, no. 13, 1993, pages 3097-3103, XP002121559 OXFORD UNIVERSITY PRESS, SURREY., GB ISSN: 0305-1048 the whole document	1
P,X	WO 98 24810 A (JANSSEN PHARMACEUTICA) 11 June 1998 (1998-06-11) page 1 -page 97 claims 1-125	1-31, 49-54, 60-67, 72-80, 85,88, 91-95
P,X	NAGASE T ET AL.: "Human mRNA for KIAA0930 protein" EMBL SEQUENCE DATABASE, 9 April 1999 (1999-04-09), XP002121417 HEIDELBERG DE cited in the application Accession Nr.: AB023155 abstract	1-11

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP 99/03848

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
see FURTHER INFORMATION sheet PCT/ISA/210

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 32-37, 40-45, 49-54, 56, 57, 68-71, 81-84, 86, 87, 89

Present claim 1 relates to an extremely large number of possible vertebrate protein homologues of a UNC-53 protein of C.elegans. Support within the meaning of Article 6 PCT and/or disclosure within the meaning of Article 5 PCT is to be found, however, for only a very small proportion of the homologues claimed, i.e. only for the human homologue hs-unc-53/3 (see description page 1, lines 31-34 and page 2, lines 12-15). In the present case, the claims so lack support, and the application so lacks disclosure, that a meaningful search over the whole of the claimed scope is impossible. Consequently, the search has been carried out for those parts of the claims which appear to be supported and disclosed, namely those parts relating to hs-unc-53.

Claims 32-37, 40-45, 49-54, 56, 57, 68-71, 81-84, 86, 87 and 89 have not been searched, because they relate to compounds (and their use) whose structural features have not been disclosed at all. Thus, these claims totally lack support within the meaning of Article 6 PCT and/or disclosure within the meaning of Article 5 PCT.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

INTERNATIONAL SEARCH REPORT

...information on patent family members

Interi ...plication No
PCT/EP 99/03848

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9638555 A	05-12-1996	AU 6123496 A	18-12-1996
		EP 0832222 A	01-04-1998
WO 9824810 A	11-06-1998	AU 5662298 A	29-06-1998
		EP 0941239 A	15-09-1999